Adaptive estimation on anisotropic Hölder spaces Part I. Fully adaptive case

Nicolas Klutchnikoff

Adaptive estimation on anisotropic Hölder spaces
Part I. Fully adaptive case

Nicolas Klutchnikoff*

April 19, 2006

Abstract

In this paper it is assumed that a noisy multidimensionnal signal is observed (for example an image in dimension 2) and our goal is to reconstruct it *as best as possible*.

In order to achieve this goal, we consider the well known theory of adaptation on a minimax sense : we want to construct a single estimator which achieves on each fuctionnal space of a given collection the "best possible rate". We introduce a new criterion in order to chose an optimal family of normalizations. This criterion is more sophisticated than criteria given by Lepski (1991) and Tsybakov (1998) and well adapted to multidimensionnal case.

Then, we prove a result of adaptation with respect to a collection of anisotropic Hölder spaces. We construct a procedure (based on comparison of kernel estimators in order to chose, according to the observations, the best among a family) which is proved to be optimal in our sense.

1 Introduction

1.1 Model

In this paper, it is supposed that we observe the "trajectory" $\mathcal{X}^{(\varepsilon)} = \{X_\varepsilon(u)\}_{u\in\mathcal{D}}$ of a noisy signal which satisfies, on $[0,1]^d$ the following SDE:

$$X_\varepsilon(du) = \mathbf{f}(u)du + \varepsilon W(du)$$

*Université de Provence, LATP, UMR CNRS 6632. Mail: klutchni@cmi.univ-mrs.fr

where $\varepsilon > 0$ is a small parameter which represents the noise level, W is a Gaussian white noise on $[0,1]^d$ and $\mathbf{f}$ is the unknown signal to be estimated. It is assumed that $\mathbf{f}$ belongs to a given functional space $\Sigma(\varkappa)$ defined by a parameter $\varkappa$ which belongs to $\mathcal{J} \subset \mathbf{R}^m$.

Let us note that all results obtained in the paper remains valid if one replaces $[0,1]^d$ by an open set in $\mathbf{R}^d$.

Further, we will consider a particular case of this general framework where the functional space $\Sigma(\varkappa)$ is an anisotropic Hölder space $H(\beta, L)$. The exact definition of this space will be done later. Here, we mention only that $\beta = (\beta_1, \ldots, \beta_d)$ is an anisotropic (different in different directions) smoothness i.e. $\beta_i > 0$ represents the smoothness of the signal in the i^{th} direction and $L > 0$ is a Lipschitz constant. In this case $\varkappa = (\beta, L)$ and $m = d + 1$.

1.2 Quality of estimation. Minimax approach

Our goal is to estimate $\mathbf{f}$ at a given point $t \in (0,1)^d$. First, let us suppose that the "nuisance" parameter $\varkappa$ is known. To measure the quality of an arbitrary estimator $\tilde{f}_\varepsilon(\cdot) = \tilde{f}(\cdot\,; \mathcal{X}^{(\varepsilon)})$, we introduce its maximal risk on $\Sigma(\varkappa)$ as follows:

$$\forall q > 0, \quad R_\varepsilon^{(q)}\left[\tilde{f}_\varepsilon, \Sigma(\varkappa)\right] = \sup_{f \in \Sigma(\varkappa)} \mathbf{E}_f \left[\left|\tilde{f}_\varepsilon(t) - f(t)\right|^q\right].$$

We are interested in finding the asymptotic of the minimax risk (minimax rate of convergence):

$$N_\varepsilon^q(\varkappa) \asymp \inf_{\tilde{f}_\varepsilon} R_\varepsilon^{(q)}\left[\tilde{f}_\varepsilon, \Sigma(\varkappa)\right],$$

where the infimum is taken over all possible estimators.

Besides the finding $N_\varepsilon(\varkappa)$, we seek an estimator $\hat{f}_\varepsilon(\cdot)$ which achieves this rate i.e.

$$R_\varepsilon^{(q)}\left[\hat{f}_\varepsilon, \Sigma(\varkappa)\right] \asymp N_\varepsilon^q(\varkappa) \tag{U.B}$$

Any such estimator is called minimax.

Let us return to the anisotropic Hölder spaces. The solution of minimax problem was found in []. The minimax rate $N_\varepsilon(\beta, L)$ is given by the formula:

$$N_\varepsilon(\beta, L) = L^{1/(2\bar{\beta}+1)} \varepsilon^{2\bar{\beta}/(2\bar{\beta}+1)} \text{ where } 1/\bar{\beta} = \sum_{i=1}^d 1/\beta_i.$$

This rate is achieved by a kernel estimator with properly chosen kernel K and bandwidth $\eta = (\eta_1, \ldots, \eta_d)$. This estimator depends explicitly on (β, L)

at least through its bandwidth defined by

$$\eta_i = \left(\frac{\varepsilon \|K\|}{L} \right)^{\frac{2\bar{\beta}}{2\bar{\beta}+1} \frac{1}{\beta_i}}.$$

This is the typical situation: the solution of the minimax problem (rate of convergence, estimator) usually depends on the space where the minimax risk is defined.

1.3 Adaptive point of view

In practice, this dependence can be awkward. For instance, it is difficult to imagine that the smoothness β is exactly known. Usually, only the information on the belonging of the nuisance parameter to some set is available.

Formally, it is supposed that $\varkappa \in \mathcal{I} \subseteq \mathcal{J}$ or, in other words, $\mathbf{f}$ belongs to a known union of functional spaces.

Of course, we can adopt the minimax strategy: $\Sigma(\mathcal{I}) = \bigcup_{\varkappa \in \mathcal{I}} \Sigma(\varkappa)$ can be viewed as a new functional space. It is clear that the minimax on $\Sigma(\mathcal{I})$ estimator is independent on $\varkappa$. Let us note, nevertheless, that the minimax rate on this space, in other words the accuracy of minimax estimator, is not better than the "worse rate" $\sup_{\varkappa \in \mathcal{I}} N_\varepsilon(\varkappa)$. Therefore, if $\{N_\varepsilon(\varkappa)\}_{\varkappa \in \mathcal{I}}$ does not depend on $\varkappa$, this approach seems to be reasonable. For example, let us consider the family of anisotropic Hölder spaces $H(\beta, L), (\beta, L)$ such that $\bar{\beta} = \gamma$, then we have $N_\varepsilon(\beta, L) \asymp \varepsilon^{2\gamma/(2\gamma+1)}$ which is independent on β. On the other hand, in general situation, it is possible that for some $\varkappa$, $N_\varepsilon(\varkappa) \nrightarrow 0, \varepsilon \to 0$ or tends to 0 very slowly (in the previous example, it corresponds to the small values of the anisotropic smoothness β). Therefore, if $N_\varepsilon(\varkappa)$ is different (in order) for different values of $\varkappa$, this approach is not satisfactory.

Thus, we still seek a single estimator, but its accuracy should depend on the nuisance parameter $\varkappa$. Evidently we would like to have an estimator "as precise as possible" for each value of $\varkappa$. It leads to the first question arising in adaptive estimation. Does whether exist a single estimator that attains the minimax rate of convergence $N_\varepsilon(\varkappa)$ simultaneously on each space $\Sigma(\varkappa)$? Such estimator, if exists, is called optimal adaptive []. Note that the accuracy given by this estimator cannot be improved for each value of $\varkappa$.

Unfortunately, optimal adaptive estimators do not always exist (Law, Tsybakov, Lepski). In this case we need a criterion of optimality in order to determine "the best estimator". To do it, we will follow the adaptive approach which consists in the following.

1. For any estimator $\tilde{f}_\varepsilon(\cdot)$, we consider the *family* of the normalized risks indexed by $\varkappa$

$$R_\varepsilon^{(q)}\left[\tilde{f}_\varepsilon, \Sigma(\varkappa), \psi_\varepsilon(\varkappa)\right] = \sup_{f\in\Sigma(\varkappa)} \mathbf{E}_f\left[\left(\psi_\varepsilon^{-1}(\varkappa)\left|\tilde{f}_\varepsilon(t) - f(t)\right|\right)^q\right], \; \varkappa \in \mathcal{I},$$

   and let the family of normalizations $\Psi = (\psi_\varepsilon(\varkappa))_{\varkappa\in\mathcal{I}}$ and the estimator $f_\varepsilon^\Psi(\cdot)$ be such that

$$\limsup_{\varepsilon\to 0} \sup_{\varkappa\in\mathcal{I}} R_\varepsilon^{(q)}\left[\tilde{f}_\varepsilon^\Psi, \Sigma(\varkappa), \psi_\varepsilon(\varkappa)\right] < +\infty. \qquad \text{(A.U.B)}$$

   The family Ψ is called admissible.

2. We propose the criterion allowing to define the best admissible family of normalizations $\Phi = (\varphi_\varepsilon(\varkappa))_{\varkappa\in\mathcal{I}}$.

3. We construct an estimator f_ε^Φ satisfying (A.U.B) with $\Psi = \Phi$. This estimator will be called adaptive estimator.

Remark 1. *The main difficulty in realization of this program consists in finding a suitable criterion of optimality.*

- *The first attempt to give a satisfactory definition was undertaken in []. Then, it has been refined in []. In spite of the fact that these criteria can be applied to any statistical model, both of them are too rough in order to treat multidimensional problems. Below we present the criterion of optimality which generalizes the previous ones.*

- *The existence of an optimal adaptive estimator means that (A.U.B) is satisfied with $\Psi = N \triangleq (N_\varepsilon(\varkappa))_{\varkappa\in\mathcal{I}}$, i.e. N is admissible. In this case, any optimality criterion should guarantee that this family is optimal because it is impossible to improve $N_\varepsilon(\varkappa)$ for all $\varkappa$.*

1.4 Our results

In the present paper we study two different problems of adaptive estimation with respect to the collection of anisotropic Hölder spaces.

First, we consider the case when the nuisance parameter $\varkappa = (\beta, L)$ is completely unknown. In this case we find the optimal family of normalizations (in view of new criterion of optimality) and construct the adaptive estimator associated with this family (satisfying (A.U.B)). In particular, our result implies

- the optimal adaptive estimator does not exist for this estimation problem;

- the optimal family of normalization differs from the family of minimax rates of convergence $N_\varepsilon(\beta, L), (\beta, L) \in \mathcal{I}$, by a $\sqrt{\ln(1/\varepsilon)}$-factor that can be viewed as the price to pay for adaptation.

In dimension 1 the similar result was obtained in [] using another criterion of optimality. We replace it by finer criterion which is more suitable for multidimensional case.

It is worth to mention that our adaptive procedure is quite different from the estimator proposed in []. As in [], our estimator is a measurable choice from the collection of the kernel estimators but the strategy of the choice is much more sophisticated due to the dimension. Similar strategy was used in [].

Let us also note that proposed method is absolutely *parameter free*, and it is applied in the situation which we treat as "fully adaptive case".

The results discussed above form the first part of this paper.

Next, we suppose that the following additional information is available. The nuisance parameter $\varkappa = (\beta, L)$ is such that $1/\bar{\beta} = 1/\gamma$ where γ is given number. Let us make several remarks:

1. In this case the minimax rate of convergence on $H(\beta, L)$ does not depend on β and given by $\varepsilon^{2\gamma/(2\gamma+1)}$. This additional information can be treated as follows. We fix the desirable accuracy of estimation (choosing parameter γ) and look for an estimator providing it for any values of nuisance parameter (β, L). The important remark, here, is that the estimator attaining the rate $\varepsilon^{2\gamma/(2\gamma+1)}$ on $H(\beta, L)$ does not achieve it on $H(\alpha, L)$ for all $\alpha \neq \beta$, $1/\bar{\alpha} = 1/\gamma$. As, minimax rates do not depend on values of β, we can adopt the minimax strategy on the union of the anisotropic Hölder spaces $H(\beta, L)$ such that $1/\bar{\beta} = 1/\gamma$.

   We show that the minimax rate is asymptotically equivalent to

$$\left(\varepsilon\sqrt{\ln\ln(1/\varepsilon)}\right)^{2\gamma/(2\gamma+1)}$$

   and construct the minimax estimator.

2. The construction of the minimax estimator (for given γ) uses the adaptive estimator obtained in the first part of the paper. This estimator could be called "partially adaptive" because the nuisance parameter (β, L) is unknown but not completely.

3. Note that found asymptotics differs from the minimax rate of convergence on $H(\beta, L), 1/\beta = 1/\gamma$ by the $\sqrt{\ln\ln(1/\varepsilon)}$-factor. It implies immediately that optimal adaptive estimators do not exist.
4. Finally, let us note that the additional information allows to minimize the price to pay for adaptation. As we mentioned before, this payment is $\sqrt{\ln(1/\varepsilon)}$ in the "fully adaptive case" and $\sqrt{\ln\ln(1/\varepsilon)}$ in the "partially adaptive case".

The partially adaptive problem forms the second part of this paper.

2 Basic definitions

2.1 Definition of the optimality

2.1.1 Motivations

As we already mentioned the first problem appearing in adaptive estimation is the existence of an optimal adaptive estimator (OAE). We will show that OAE with respect to the family of anisotropic Hölder spaces $\{H(\beta, L)\}_{(\beta,L)}$ does not exist. More precisely, we show that, for any estimator $\tilde{f}_\varepsilon$ there exists a value of the nuisance parameter, $(\beta_0, L_0) \in \mathfrak{J}$:

$$\limsup_{\varepsilon\to 0} \sup_{f\in H(\beta_0,L_0)} \mathbf{E}_f \left[\left(\varepsilon^{-\frac{2\bar{\beta}_0}{2\bar{\beta}_0+1}} \left| \tilde{f}_\varepsilon(t) - f(t) \right| \right)^q \right] = +\infty$$

Formally, this result means that the family of rates of convergence is not admissible.

In general case the non existence of an OAE can be formulated as follows: for any admissible family Ψ, there exists a nuisance parameter $\varkappa_0$ such that

$$\frac{\psi_\varepsilon(\varkappa_0)}{N_\varepsilon(\varkappa_0)} \xrightarrow[\varepsilon\to 0]{} +\infty. \tag{1}$$

Clearly, it implies that any admissible family Ψ can be "improved", at least, in this point. In particular, one can use a minimax on $\Sigma(\varkappa_0)$ estimator for all values of nuisance parameter $\varkappa$. This estimator, which could be very bad for all $\varkappa \neq \varkappa_0$, would outperform any estimator satisfying (AUB) with Ψ verifying (1).

As we see, the set of points where an admissible family can be improved is non empty. In this in mind, we will use the following principle in order to give the notion of optimality:

the "best admissible family" of normalizations should have "small number of points" where it can be improved.

2.1.2 Definition

Now, let us consider a general statistical experience $(V^\varepsilon, \mathcal{A}^\varepsilon, \{\mathbf{P}_f^\varepsilon\}_{f\in\Sigma})$ generated by the observation $\mathcal{X}^{(\varepsilon)}$, and let us suppose that $\Sigma = \bigcup_{\varkappa\in\mathcal{J}} \Sigma(\varkappa)$ where $\mathcal{J} \subset \mathbf{R}^m$ $(m \geq 1)$.

The goal is to estimate a functional $G(f)$ where $G : \Sigma \to (\Lambda, \|\cdot\|)$ where $(\Lambda, \|\cdot\|)$ is a Banach space.

In our particular case, let us recall that $(\Lambda, \|\cdot\|) = (\mathbf{R}, |\cdot|)$ and $G(f) = f(t)$. The maximal risk of an estimator $\tilde{f}_\varepsilon(\cdot)$ over the class $\Sigma(\varkappa)$ normalized by $\psi_\varepsilon(\varkappa)$ is defined by the formula

$$R_\varepsilon^{(q)}\left(\tilde{f}_\varepsilon, \Sigma(\varkappa), \psi_\varepsilon(\varkappa)\right) = \sup_{f\in\Sigma(\varkappa)} \mathbf{E}_f\left[\left(\psi_\varepsilon^{-1}(\varkappa)\|G(\tilde{f}_\varepsilon) - G(f)\|\right)^q\right].$$

This risk can be defined with a general loss function w satisfying usual assumptions and such that $w(u) \to +\infty, u \to +\infty$.

Let us introduce some definitions.

A family of normalizations $\Psi = (\psi_\varepsilon(\varkappa))_{\varkappa\in\mathcal{I}}$ is called admissible if there exists an estimator f_ε^Ψ such that the following inequality holds:

$$\limsup_{\varepsilon\to 0} \sup_{\varkappa\in\mathcal{I}} R_\varepsilon^{(q)}\left(f_\varepsilon^\psi, \Sigma(\varkappa), \psi_\varepsilon(\varkappa)\right) < +\infty.$$

For two admissible families Ψ and $\tilde{\Psi}$, we introduce two sets

$$\begin{aligned}
\mathcal{I}_0(\Psi/\tilde{\Psi}) &= \left\{\kappa \in \mathcal{I} : \quad \frac{\psi_\varepsilon(\kappa)}{\tilde{\psi}_\varepsilon(\kappa)} \xrightarrow[\varepsilon\to 0]{} 0\right\}; \\
\mathcal{I}_\infty(\Psi/\tilde{\Psi}) &= \left\{\varkappa \in \mathcal{I} : \quad \frac{\psi_\varepsilon(\kappa)}{\tilde{\psi}_\varepsilon(\kappa)} \times \frac{\psi_\varepsilon(\varkappa)}{\tilde{\psi}_\varepsilon(\varkappa)} \xrightarrow[\varepsilon\to 0]{} +\infty, \forall\kappa \in \mathcal{I}_0(\Psi/\tilde{\Psi})\right\}.
\end{aligned}$$

The set $\mathcal{I}_0(\Psi/\tilde{\Psi})$ consists of all points where Ψ is "better", in order, than $\tilde{\Psi}$. One can say that $\tilde{\Psi}$ is "dominated" by Ψ on $\mathcal{I}_0(\Psi/\tilde{\Psi})$.

On the contrary, the set $\mathcal{I}_\infty(\Psi/\tilde{\Psi})$ consists of the points where $\tilde{\Psi}$ "dominates" Ψ and, moreover, the loss of $\tilde{\Psi}$ w.r.t. Ψ on $\mathcal{I}_0(\Psi/\tilde{\Psi})$ is "compensated".

Our principle of the choice between two admissible families Ψ and $\tilde{\Psi}$ consists in comparing of "massivities" of $\mathcal{I}_0(\Psi/\tilde{\Psi})$ and $\mathcal{I}_\infty(\Psi/\tilde{\Psi})$:

$\tilde{\Psi}$ is "better" than Ψ if $\mathcal{I}_\infty(\Psi/\tilde{\Psi})$ is much more "massive" than $\mathcal{I}_0(\Psi/\tilde{\Psi})$.

This idea leads to the following definition of an optimal family of normalizations.

Not to give the additional definitions, here and later, we will suppose that $\mathcal{I}$ contains an open set of $\mathbf{R}^m$.

Definition 1. *I) A family of normalizations $\Phi = (\varphi_\varepsilon(\varkappa))_{\varkappa \in \mathcal{I}}$ is called "optimal" if:*

i) Φ is an "admissible" family.

ii) If $\Psi = (\psi_\varepsilon(\varkappa))_{\varkappa \in \mathcal{I}}$ is another admissible family of normalizations we have:

- *$\mathcal{I}_0(\Psi/\Phi)$ is contained in a $(m-1)$-manyfold,*
- *$\mathcal{I}_\infty(\Psi/\Phi)$ contains an open set of $\mathbf{R}^m$.*

II) The estimator $f_\varepsilon^\Phi(\cdot)$ is called an adaptive estimator.

Let us comment this criterion.

Remark 2. *1. This definition is correct in the following sense:*

- *if Φ and $\tilde{\Phi}$ are two optimal families, then:*

$$\varphi_\varepsilon(\varkappa) \asymp \tilde{\varphi}_\varepsilon(\varkappa),\ \forall \varkappa \in \mathcal{I}.$$

- *If N is an admissible family (i.e. there exists an OAE), then it satisfies Definition 1. Indeed, in this case, $\mathcal{I}_0(\Psi/N)$ is empty, for any Ψ.*

*2. Note that we well followed our principle: the set of points where the optimal family Φ can be improved is really "small". Indeed the "dimension" of the set $\mathcal{I}_0(\Psi/\Phi)$ is **strictly** less than the dimension of $\mathcal{I}$ for **any** Ψ.*

3. Let us also note that, the non existence of OAE implies that there exists Ψ such that $\mathcal{I}_0(\Psi/\Phi) \neq \emptyset$ i.e. there exists normalization (may be not unique) which "dominates" Φ on $\mathcal{I}_0(\cdot/\Phi)$. Let us denote $\mathfrak{N}$ the set of all normaliszations dominating Φ.

The message we would like to address is that the estimator f_ε^Φ (satisfying (AUB) with Φ) "outperforms" any estimator f_ε^Ψ satisfying (AUB) with Ψ belonging to $\mathfrak{N}$.

Indeed, the estimator f_ε^Ψ is more precise than f_ε^Φ on $\mathcal{I}_0(\Psi/\Phi)$, which, let us remind is "very small set" for any $\Psi \in \mathfrak{N}$. The loss of f_ε^Φ w.r.t. f_ε^Ψ is given by

$$\{\varphi_\varepsilon(\varkappa)/\psi_\varepsilon(\varkappa) : \varkappa \in \mathcal{I}_0(\Psi/\Phi)\}.$$

On the other hand, the estimator f_ε^Φ is more precise than f_ε^Ψ at least on $\mathcal{I}_\infty(\Psi/\Phi)$, which, in view of Definition 1, is very large. The gain of f_ε^Φ w.r.t. f_ε^Ψ is given, at least, by

$$\{\varphi_\varepsilon(\varkappa)/\psi_\varepsilon(\varkappa) : \varkappa \in \mathcal{I}_\infty(\Psi/\Phi)\}.$$

*In view of the definition of $\mathcal{I}_\infty(\cdot/\Phi)$, we can conclude that the gain of f_ε^Φ w.r.t. f_ε^Ψ (for **any** $\Psi \in \mathfrak{N}$!), is much* bigger *on the* larger *set than its* loss *on the* smaller *set.*

2.2 Anisotropic Hölder spaces

To define the class of Hölder spaces let us introduce some notations. A function f belongs to $\mathcal{C}_\mathcal{D}$ if f is from $\mathbf{R}^d$ to $\mathbf{R}$ and it is compactly supported on $\mathcal{D}$. For a such function f, $i \in [\![1;d]\!]$ and $x \in \mathbf{R}^d$ we define:

$$\begin{array}{rcl} f_i(\cdot|x) : \mathbf{R} & \to & \mathbf{R} \\ y & \mapsto & f(x_1, \ldots, x_{i-1}, x_i + y, x_{i+1}, \ldots, x_d) \end{array}$$

Let us denote $m_i(\beta) = \sup\{n \in \mathbf{N}; n < \beta_i\}$ and $\alpha_i(\beta) = \beta_i - m_i(\beta)$.

Definition 2. *Set $(\beta, L) \in \mathfrak{J}$. A function $f \in \mathcal{C}_\mathcal{D}$ belongs to the anisotropic Hölder space $H(\beta, L)$ if:*

- *The following property holds:*

$$\sup_{i=1,\ldots,n} \sup_{x \in \mathbf{R}^d} \sum_{s=0}^{m_i} \left\| f_i^{(s)}(\cdot|x) \right\|_\infty \leq L,$$

- *for all $y \in \mathbf{R}$ and all $i \in [\![1;d]\!]$, the following inequality holds:*

$$\sup_{x \in \mathbf{R}^d} \left| f_i^{(m_i)}(y|x) - f_i^{(m_i)}(0|x) \right| \leq L|y|^{\alpha_i},$$

where $m_i = m_i(\beta)$ and $\alpha_i = \alpha_i(\beta)$.

In words, on the i^{th} direction of the canonical base of $\mathbf{R}^d$ the Hölder regularity (in the classical sense) of f is (β_i, L).

3 Our goal

Here and later, we consider the "fully adaptive problem".

Set $b = (b_1, \ldots, b_d) \in (\mathbf{R}_+^*)^d$ and $l_* > 0$. Let us define, for all (β, L) such that $\bar{\beta} \leq \bar{b}$ and $L \geq l_*$, the following quantities:

$$\rho_\varepsilon(\beta, L) = \sqrt{1 + \frac{4(\bar{b} - \bar{\beta})}{(2\bar{b}+1)(2\bar{\beta}+1)} \ln \frac{L}{\|K\|\varepsilon} + \frac{2}{2\bar{b}+1} \ln \frac{L}{l_*}}$$

and

$$\varphi_\varepsilon(\beta, L) = L^{1/(2\bar{\beta}+1)} \left(\|K\|\varepsilon\rho_\varepsilon(\beta, L)\right)^{2\bar{\beta}/(2\bar{\beta}+1)}.$$

Here and later, it is assumed that $\varepsilon < l_*/\|K\|$. This assumption guarantees that $\rho_\varepsilon(\beta, L)$ (which will be viewed as the price to pay for adaptation) is greater than 1.

Let us denote $\Phi = (\varphi_\varepsilon(\beta, L))$. Our goal is to prove that Φ is the optimal family of normalizations w.r.t our criterion and, moreover, to construct an adaptive estimator, namely $f_\varepsilon^\Phi(\cdot)$, which satisfies (A.U.B) with Φ.

Remark 3. *At point (b, l_*), it is impossible to improve φ_ε since it corresponds at the minimax rate of convergence $N_\varepsilon(b, l^*)$. Let us also note that:*

$$\rho_\varepsilon(\beta, L) = \sqrt{1 + 2\ln \frac{N_\varepsilon(\beta, L)}{N_\varepsilon(b, l_*)}}$$

4 Adaptive procedure

In this section, we describe the adaptive procedure. Let us recall that this procedure is constructed by the choice (data dependent) from the collection of† *kernel estimators.*

4.1 Kernels

A kernel is a function from $\mathbf{R}^d$ to $\mathbf{R}$ with some additional properties. We will denote $\mathcal{K}$ the class of kernel we will use. A kernel K belongs to $\mathcal{K}$ if it belongs to $\mathbf{L}^1(\mathbf{R}^d) \cap \mathbf{L}^2(\mathbf{R}^d)$ and satisfies the following properties:

$$\int_{\mathbf{R}^d} K(u)du = 1 \tag{K1}$$

$$\forall i \in [\![1; d]\!], \quad \int_{\mathbf{R}^d} |K(u)|(1 + |u_i|)^{b_i} du < +\infty \tag{K2}$$

$$\forall i \in [\![1; d]\!], \quad \int_{[-1,1]^d} |K(u)| du > 0. \tag{K3}$$

$$\forall i \in [\![1; d]\!], \forall l \in [\![1; b_i]\!], \quad \int_{\mathbf{R}^d} K(u) u_i^l du = 0. \tag{K4}$$

Further, "K is a kernel" will signify "K is a kernel belonging to $\mathcal{K}$". Then we will denote

$$\|K\| = \left(\int_{\mathbf{R}^d} |K(u)|^2 du \right)^{1/2}.$$

Remark 4. *Condition (K3) is a technical one. Other assumptions are classical.*

4.2 Collection of kernel estimators

First, for each $k \in \mathbf{Z}^d$ we define a bandwidth $h^{(k)} = (h_1^{(k)}, \dots, h_d^{(k)})$ in the following way. We introduce

$$h(b, l_*, \varepsilon) = \left(\frac{\|K\|\varepsilon}{l_*} \right)^{\frac{2\bar{b}}{2\bar{b}+1} \frac{1}{b_i}},$$

and, therefore

$$\forall i \in [\![1; d]\!], \quad h_i^{(k)} = h(b, l_*, \varepsilon) 2^{-(k_i+1)}.$$

Then, we can introduce, for all $k \in \mathbf{Z}^d$ the normalized kernel

$$K_k(u) = \left(\prod_{i=1}^d h_i^{(k)} \right)^{-1} K\left(\frac{u_1}{h_1^{(k)}}, \dots, \frac{u_d}{h_d^{(k)}} \right),$$

and the associated kernel estimator:

$$\hat{f}_k(t) = \int_{\mathbf{R}^d} K_k(t-u) X_\varepsilon(du).$$

Then, let us define

$$N_\varepsilon = \left\lfloor 2 \left(\frac{2\bar{b}}{2\bar{b}+1} \ln \frac{l_*}{\|K\|\varepsilon} + \ln \frac{l^*}{l_*} \right) \right\rfloor + 1$$

and

$$C(b) = \frac{2\bar{b}+1}{2\bar{b}} \times \frac{\ln 2 + \sqrt{2 \ln 2}}{\ln 2}.$$

For all $n \in [\![0; N_\varepsilon]\!]$, we consider the set

$$\mathcal{Z}(n) = \left\{ k = \in \mathbf{Z}^d : \sum_{i=1}^{d} (k_i + 1) = n \text{ and } \forall i, |k_i| \leq C(b)n + 1 \right\} \tag{2}$$

where $k = (k_1, \ldots, k_d)$. This enables us to define

$$\mathcal{Z}_\varepsilon = \bigcup_{n=0}^{N_\varepsilon} \mathcal{Z}(n).$$

Finaly, we define the collection of estimators $\{\hat{f}_k(\cdot)\}_{k \in \mathcal{Z}_\varepsilon}$.

4.3 Procedure

4.3.1 Useful notations

Set

$$\begin{cases} \lambda_* &= \min_{i \in [\![1;d]\!]} \int_{[-1,1]^d} |K(u)| \frac{|u_i|^{b_i}}{m_i(b)!} du \\ \lambda^* &= \max_{i \in [\![1;d]\!]} \int_{\mathbf{R}^d} |K(u)| (1 + |u_i|)^{b_i} \end{cases}$$

and let

$$C = \frac{4}{\lambda_* d} + 2\sqrt{6q + 4}.$$

Let us define a "partial ordering" on $\mathcal{Z}_\varepsilon$. We say that $k \preceq l$ if:

$$\sum_{i=1}^{d} (k_i + 1) \triangleq |k| \leq |l| \triangleq \sum_{i=1}^{d} (l_i + 1).$$

Let us also define, for all k and l in $\mathcal{Z}_\varepsilon$, $k \wedge l \in \mathbf{Z}^d$ by the formula:

$$k \wedge l = (k_1 \wedge l_1, \ldots, k_d \wedge l_d).$$

The following quantities will be used throughout this paper.

$$\sigma_\varepsilon(l) = \frac{\varepsilon \|K\|}{\left(\prod_{i=1}^{d} h_i^{(l)} \right)^{1/2}}, \tag{3}$$

and

$$S_\varepsilon(l) = \sigma_\varepsilon(l) \sqrt{1 + |l| \ln 2}.$$

Remark 5. *Let us note that:*

$$\sigma_\varepsilon(l) = \sqrt{\mathrm{Var}\left(\hat{f}_l\right)}.$$

4.3.2 Adaptive estimator

Now, let us explain how the procedure chose an estimator. We introduce the random set $\mathcal{A}$ of all indexes defined by:

$$\mathcal{A} = \left\{ k \in \mathcal{Z}_\varepsilon : \left| \hat{f}_{k\wedge l}(t) - \hat{f}_l(t) \right| \leq C S_\varepsilon(l), \; \forall l \in \mathcal{Z}_\varepsilon, \, l \succeq k \right\}.$$

If $\mathcal{A}$ is non empty, one can chose $\hat{k}$ (may be not unique) such that:

$$\hat{k} = \arg\min_{k \in \mathcal{A}} \sigma_\varepsilon(k).$$

Remark that this choice of $\hat{k}$ is measurable because $\mathcal{A}$ is a finite set ($\mathcal{A} \subseteq \mathcal{Z}_\varepsilon$, $\mathcal{Z}_\varepsilon$ finite). Now, we can construct explicitly our estimator f_ε^Φ in the following way:

$$f_\varepsilon^\Phi(t) = \begin{cases} \hat{f}_{\hat{k}}(t) & \text{if } \mathcal{A} \neq \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

4.4 Comments

Let us make several comments about this procedure.

- Our procedure is quite different to that introduced in [] to solve the similar one-dimensional problem. The main difference is connected with the manner of choosing a random index $\hat{k}$. Here, we compare *pairwise* the estimators by introducing the "artificial" estimator $\hat{f}_{k\wedge l}(\cdot)$ in definition of $\mathcal{A}$. Then, we chose an estimator of minimal variance estimators $\{\hat{f}_k\}_{k \in \mathcal{A}}$.

  In dimension 1, it is useless to introduce these artificial estimators. Unfortunately, a such procedure is not adapted to solve multidimensional problems (if one deals with anisotropic regularities) and fails.

- Our procedure is inspired by the method proposed in [], well adapted to multidimensional problems. Let us mention however, that it is not possible to use this method directly. The main difference is the choice of set of indexes. In our case, we have to consider (it will be explained further in this paper) indexes belonging to $\mathbf{Z}^d$ — instead of $\mathbf{N}^d$ considered in []. First of all, let us remind that our set of indexes is

$$\mathcal{Z}_\varepsilon = \bigcup_{n=0}^{N_\varepsilon} \mathcal{Z}(n) \subseteq \mathbf{Z}^d.$$

The set used in [] is $\mathcal{N}_\varepsilon = \bigcup_n \mathcal{N}(n)$ where:

$$\mathcal{N}(n) = \left\{ k = (k_1, \ldots, k_d) \in \mathbf{N}^d : \sum_{i=1}^{d} (k_i + 1) = n \right\}.$$

Both procedures require the following properties of indexes:

$$\left\{ \begin{array}{rcl} \displaystyle\sum_{n=0}^{\infty} 2^{-n} \# \mathcal{Z}(n) & < & \infty \\ \displaystyle\sum_{n=0}^{\infty} 2^{-n} \# \mathcal{N}(n) & < & \infty. \end{array} \right. \tag{4}$$

Second property is evidently fulfilled. The first one requires a special construction given by (2).

Let us also note that not to pay an additional price at final point (b, l_*), we need $\mathcal{Z}(0)$ is bounded independently on ε which follows immediately from 4 otherwise we need to pay $\sqrt{1 + \ln \mathcal{Z}(0)}$.

Remark 6. *Figure 1 represents $\mathcal{Z}_\varepsilon$ in dimension 2. Here, $\mathcal{Z}(i), i = 0, 1, 2$ and $\mathcal{Z}(n+2)$ are drawn.The black points belong to $\mathcal{Z}_\varepsilon$.*

4.5 Upper bound

Let us introduce some basic notations: $b = (b_1, \ldots, b_d)$ is a vector of positive numbers and $0 < l_* < l^* < +\infty$ are given. Set

$$\mathcal{B} = \prod_{i=1}^{d} (0, b_i] \text{ and } \mathcal{L} = [l_*, l^*].$$

Moreover, for all $\gamma \in (0, \bar{b}]$, let us consider

$$\mathcal{B}(\gamma) = \left\{ \beta \in \mathcal{B}; 1/\bar{\beta} = 1/\gamma \right\}.$$

Let us denote

$$\mathfrak{J} = \mathcal{B} \times \mathcal{L} \text{ and } \mathfrak{J}(\gamma) = \mathcal{B}(\gamma) \times \mathcal{L}.$$

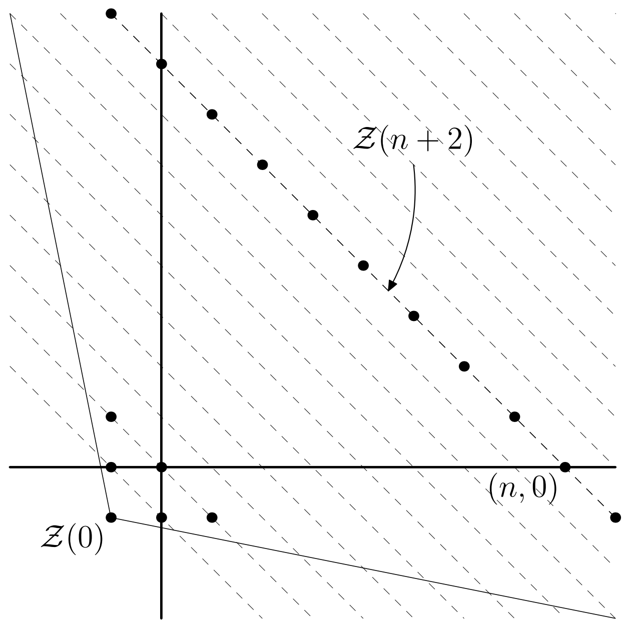
Theorem 1. *Set $\varepsilon < l_*/\|K\|$ and $q > 0$. Then*

$$\sup_{(\beta, L) \in \mathfrak{J}} R_\varepsilon^{(q)} \left(f_\varepsilon^\Phi(t), H(\beta, L), \varphi_\varepsilon(\beta, L) \right) \leq M_q$$

where M_q is an absolute constant which does not depend on ε. The explicit expression of $M_q = M_q(l_*, l^*, b)$ is given in the proof.

Remark 7. *This result implies clearly that f_ε^Φ satisfies (A.U.B) with Φ but it is stronger: in fact, we obtain a non-asymptotical upper bound for all ε small enough.*

Figure 1: $\mathcal{Z}_\varepsilon$.



5 Optimality of Φ

5.1 Result

Theorem 2. *Set* $\psi = (\psi_\varepsilon(\beta, L))_{(\beta,L)\in\mathfrak{I}}$ *an admissible family of normalizations such that* $\exists \beta_0 \in \mathcal{I}_0(\Psi/\Phi)$*, then:*

1. $\mathcal{I}_0(\Psi/\Phi) \subseteq \mathfrak{I}(\bar{\beta}_0)$;
2. $\mathcal{I}_\infty(\Psi/\Phi) \supseteq \bigcup_{\gamma>\bar{\beta}_0} \mathfrak{I}(\gamma)$.

Remark 8. *This result implies that* Φ *is the optimal family w.r.t to our criterion. Indeed,* $\dim(\mathfrak{I}(\bar{\beta}_0)) = d < d+1 = \dim(\mathfrak{I})$ *and it is clear that* $\bigcup_{\gamma>\bar{\beta}_0} \mathfrak{I}(\gamma)$ *contains an open set of* $\mathfrak{I}$*.*

5.2 Comments

Let us briefly discuss an interesting point which "shows" that our criterion of optimality is well adapted to our problem. One of our idea, by introducing this criterion, was to minimize $\mathcal{I}_0(\Psi/\Phi)$ (in term of massivity). Theorem 2 says that this set is always contained in $\mathfrak{I}(\gamma)$ for a given γ. Can we improve this result (by proving that $\mathcal{I}_0(\Psi/\Phi)$ is essentially smaller than $\mathfrak{I}(\gamma)$)? The answer is no!

Indeed, let us suppose that the result concerning the "partially adaptive problem" is proved. Thus, for all $0 < \gamma < \bar{b}$, the minimax rate of convergence on

$$\mathcal{F}(\gamma) = \bigcup_{(\beta,L)\in\mathfrak{I}(\gamma)} H(\beta, L)$$

is given by

$$\phi_\varepsilon(\gamma) \asymp \left(\varepsilon\sqrt{\ln\ln\frac{1}{\varepsilon}}\right)^{\frac{2\gamma}{2\gamma+1}}.$$

It is evident that any estimator which achieves this rate on $\mathcal{F}(\gamma)$ outperform f_ε^Φ at least on $\mathfrak{I}(\gamma)$. The loss is about:

$$\left(\frac{\ln 1/\varepsilon}{\ln\ln 1/\varepsilon}\right)^{\frac{\gamma}{2\gamma+1}}$$

Combining this result with Theorem 2, we obtain

$$\mathcal{I}_0(\Psi^\gamma/\Phi) = \mathfrak{I}(\gamma).$$

where $\Psi^\gamma = (\psi_\varepsilon^\gamma(\beta, L))_{(\beta,L)}$ is defined by

$$\psi_\varepsilon^\gamma(\beta, L) = \begin{cases} \phi_\varepsilon(\gamma) & \text{if } \bar{\beta} = \gamma \\ 1 & \text{otherwise.} \end{cases}$$

6 Proof of theorem 1

6.1 Introduction

Let us explain, briefly, the main ideas to prove our result.

First, let us suppose that the smoothness parameter (β, L) of the signal (to be estimated) is well known. Thus, it is easy to construct an estimator (depending on (β, L)) witch achieves the expected rate $\varphi_\varepsilon(\beta, L)$.

To do that, we have to chose $\tilde{h}(\beta, L, \varepsilon) = (\tilde{h}_1(\beta, L, \varepsilon), \dots, \tilde{h}_d(\beta, L, \varepsilon))$ (bandwidth of this estimator) on the following way:

$$\tilde{h}_i(\beta, L, \varepsilon) = \gamma_i(\beta)\left(\frac{\|K\|\Gamma(\beta)}{2L}\varepsilon\rho_\varepsilon(\beta, L)\right)^{\frac{2\bar{\beta}}{2\bar{\beta}+1}\frac{1}{\beta_i}}, \forall i,$$

where

$$\begin{cases} \gamma_i(\beta) &= (\lambda_i(\beta)\beta_i)^{-1/\beta_i} \\ \Gamma(\beta) &= \left(\prod_{i=1}^{d} \gamma_i(\beta)\right)^{-1/2} \end{cases}.$$

This formula is obtained as the solution of the following minimization problem:

$$\tilde{h}(\beta, L, \varepsilon) = \arg\min_{h \in \mathcal{H}} \left(b^{\beta,L} + s_\varepsilon^{\beta,L}\right)(h) \tag{5}$$

where

$$b^{\beta,L}(h) = L \sum_{i=1}^{d} \lambda_i(\beta) h_i^{\beta_i}$$

is a *bias* term and

$$s_\varepsilon^{\beta,L}(h) = \frac{\|K\|\varepsilon}{\sqrt{\prod_{i=1}^{d} h_i}} \rho_\varepsilon(\beta, L)$$

can be viewed as a *penalized standard deviation* term.

Remark 9. *Using these notations we obtain*

$$b^{\beta,L}(\tilde{h}(\beta, L, \varepsilon)) \asymp s_\varepsilon^{\beta,L}(\tilde{h}(\beta, L, \varepsilon)) \asymp \varphi_\varepsilon(\beta, L), \quad \forall (\beta, L).$$

Next, if (β, L) is unknown, we want that our procedure choses a kernel estimator as good as the optimal one, constructed using bandwidth $\tilde{h}(\beta, L, \varepsilon)$. In order to do that, our procedure compare a large number of estimators. In particular, for each $(\beta, L) \in \mathfrak{J}$, the estimator constructed using bandwidth $\tilde{h}(\beta, L, \varepsilon)$ should be "viewed" by the procedure. This implies that set $\mathcal{Z}_\varepsilon$ is large enough.

6.2 Lemmas

Here, we give some lemmas witch will be proved in Appendix. They are used further in the proof.

Lemma 1. *Set* (β, L).

Bandwidth $\tilde{h}(\beta, L, \varepsilon)$ *is the unique bandwidth* $\tilde{\eta}$ *such that:*

$$\tilde{\eta} = \arg\min_{h \in \mathcal{H}} \left(b^{\beta,L} + s_\varepsilon^{\beta,L}\right)(h).$$

For simplicity, let us denote $h(\beta, L, \varepsilon) = (h_1(\beta, L, \varepsilon), \ldots, h_d(\beta, L, \varepsilon))$ defined by:

$$h_i(\beta, L, \varepsilon) = \left(\frac{\|K\|}{L} \varepsilon \rho_\varepsilon(\beta, L) \right)^{\frac{2\bar{\beta}}{2\bar{\beta}+1} \frac{1}{\beta_i}}.$$

It is clear that, the estimator defined using bandwidth $h(\beta, L, \varepsilon)$ is asymptoticaly as good as the estimator defined using bandwidth $\tilde{h}(\beta, L, \varepsilon)$. Indeed, for all (β, L) we have:

$$h_i(\beta, L, \varepsilon) \asymp \tilde{h}_i(\beta, L, \varepsilon), \quad \forall i.$$

Now, let us consider

$$k_i(\beta, L, \varepsilon) = \left\lfloor \frac{1}{\ln 2} \ln \left(\frac{h_i^*(\varepsilon)}{h_i(\beta, L, \varepsilon)} \right) \right\rfloor$$

where $\lfloor x \rfloor = \sup\{n \in \mathbf{N} : n \leq x\}$. And let us consider the index $k(\beta, L, \varepsilon) = (k_1(\beta, L, \varepsilon), \ldots, k_d(\beta, L, \varepsilon))$ in $\mathbf{Z}^d$.

It is easy to see that the estimator defined by the bandwidth $h^{(k(\beta,L,\varepsilon))}$ is asymptoticaly as good as the estimator defined by $h(\beta, L, \varepsilon)$ and, thus, as good as that one defined by $\tilde{h}(\beta, L, \varepsilon)$.

Lemma 2. *Set (β, L). Index $k(\beta, L, \varepsilon)$ belongs to $\mathcal{Z}_\varepsilon$.*

Remark 10. *Set $\mathcal{Z}_\varepsilon$ was constructed such that this lemma is satisfied and moreover such that inequality (4) holds.*

Let us give an improtant lemma concerning the canonical decomposition of the estimator $\hat{f}_k$.

Lemma 3. *Let us fix $f \in \bigcup_{(\beta,L)\in\mathcal{B}\times\mathcal{I}} H(\beta, L)$, and let us calculate under the law $\mathbf{P}_f$. We have, for $k \in \mathcal{Z}_\varepsilon$:*

$$\hat{f}_k(t) = f(t) + b_k(t) + \sigma_\varepsilon(k)\xi(k),$$

where:

$$\begin{cases} b_k(t) &= \displaystyle\int_{\mathbf{R}^d} K(u) \left(f(t - h^{(k)}.u) - f(t) \right) du \\ \sigma_\varepsilon(k) &= \dfrac{\|K\|\varepsilon}{\left(\prod_{i=1}^d h_i^*(\varepsilon)\right)^{1/2}} 2^{\frac{|k|}{2}} \\ \xi(k) &\sim \mathcal{N}(0,1), \end{cases}$$

where $h.u$ denotes the following vector: $(h_1u_1, \ldots, h_du_d)$. Moreover, let us remark that, if $k \preceq l$, then $\sigma_\varepsilon(k) \leq \sigma_\varepsilon(l)$.

Now, let us give the most important lemma about the control of bias terms. More precisely:

Lemma 4. *Set $(\beta, L) \in \mathcal{B} \times \mathcal{I}$ and $f \in H(\beta, L)$. Under $\mathbf{P}_f$ we have:*

$$\forall k \in \mathcal{Z}_\varepsilon, \quad |b_k(t)| \leq B^{\beta,L}(k)$$

and

$$\forall (k,l) \in \mathcal{Z}_\varepsilon^2, \quad |b_{k \wedge l}(t) - b_l(t)| \leq 2B^{\beta,L}(k),$$

where $B^{\beta,L}(k) = b^{\beta,L}(h^{(k)})$.

Now, let us give a lemma concerning the link beetwen the bias and the penalized standard deviation of the estiamtor $\hat{f}_{k(\beta,L,\varepsilon)}$. First of all let us recall that $S_\varepsilon(k)$ was define by equation (3).

Lemma 5. *For all $(\beta, L) \in \mathcal{B} \times \mathcal{I}$ we have:*

$$B^{\beta,L}(k(\beta, L, \varepsilon)) \leq C^* S_\varepsilon(k(\beta, L, \varepsilon)),$$

where $C^* = (d\lambda^*)\sqrt{2 \vee (\bar{b}+1)/\bar{b}}$.

Finally, let us give a lemma which explain a link beetwen $S_\varepsilon(k(\beta, L, \varepsilon))$ and the rate of convergence $\varphi_\varepsilon(\beta, L)$. It is very important, because this lemma proves that it is enough to control (up to a constant) the quality of the estimator f_ε^Φ by $S_\varepsilon(k(\beta, L, \varepsilon))$.

Lemma 6. *For all $(\beta, L) \in \mathcal{B} \times \mathcal{I}$, we have:*

$$S_\varepsilon(k(\beta, L, \varepsilon)) \leq \varphi_\varepsilon(\beta, L).$$

Lemma 3 is evident. All the others will be proved in Apendix.

6.3 Proof

Let us consider $\varepsilon < l_*/\|K\|$ and $q > 0$.

We want to prove that

$$\sup_{(\beta,L)\in\mathfrak{J}} \sup_{f \in H(\beta,L)} \mathbf{E}_f\left[\left(\varphi_\varepsilon^{-1}(\beta, L)\left|f_\varepsilon^\Phi(t) - f(t)\right|\right)^q\right] < +\infty.$$

Thus, we fix $(\beta, L) \in \mathfrak{J}$ and $f \in H(\beta, L)$. Let us denote $\kappa = k(\beta, L, \varepsilon)$. Let us recall that $k(\beta, L, \varepsilon)$ is the index correponding to the bandwidth $h(\beta, L, \varepsilon)$.

Now, we have to distinguish two cases: First $\mathcal{A}$ is empty. Next, it is non empty.

A) $\mathcal{A}$ is non empty

In this case, the procedure chose an index $\hat{k}$. Main idea is the following: we have to compare $\hat{f}_{\hat{k}}$ and $\hat{f}_\kappa$. To do that, let us introduce the following sets, for all $s \in \mathbf{N}$:

$$\begin{cases} B_1(s) &= \{k \in \mathcal{Z}_\varepsilon : |k| \leq |\kappa| + sd\} \\ B_2(s) &= \{k \in \mathcal{Z}_\varepsilon : |k| > |\kappa| + (s-1)d\} \end{cases}$$

Using these notations we obtain:

$$\mathcal{Z}_\varepsilon = B_1(0) \cup \left(\bigcup_{s \geq 1} B_1(s) \cap B_2(s) \right).$$

Thus, we have:

$$\begin{aligned} \mathbf{E}_f\left[\left(\left|f_\varepsilon^\Phi(t) - f(t)\right|\right)^q\right] &\leq \mathbf{E}_f\left[\left|\hat{f}_{\hat{k}}(t) - f(t)\right|^q \mathbf{1}_{\{\hat{k} \in B_1(0)\}}\right] \\ &\quad + \sum_{s \geq 1} \mathbf{E}_f\left[\left|\hat{f}_{\hat{k}}(t) - f(t)\right|^q \mathbf{1}_{\{\hat{k} \in B_1(s)\}} \mathbf{1}_{\{\hat{k} \in B_2(s)\}}\right] \\ &\leq R(0,q) + \sum_{s \geq 1} \sqrt{R(s,2q)D(s)}, \end{aligned}$$

where

$$\begin{cases} R(s,p) = \mathbf{E}_f\left[\left|\hat{f}_{\hat{k}}(t) - f(t)\right|^p \mathbf{1}_{\{\hat{k} \in B_1(s)\}}\right] \\ D(s) = \mathbf{P}_f\left[\hat{k} \in B_2(s)\right]. \end{cases}$$

Thus we have to control these quantities.

Control of $\boldsymbol{D(s)}$. Let us denote $\kappa(s) = (\kappa_1 + s, \ldots, \kappa_d + s)$ and let us consider $s_{\max} = \max\{s \in \mathbf{N} : B_2(s) \neq \emptyset\}$. Clearly $s_{\max} \leq N_\varepsilon + 1$. Thus we have $D(s) = 0$ for any $s > s_{\max}$. Let us consider $s \leq s_{\max}$. It is easy to see that $\hat{k} \in B_2(s) \Rightarrow \kappa(s-1) \notin \mathcal{A}$. Thus, if $\hat{k} \in B_2(s)$, then there exists $l \in \mathcal{Z}_\varepsilon, l \succeq \kappa(s-1)$, such that

$$\left|\hat{f}_{\kappa(s-1)\wedge l}(t) - \hat{f}_l(t)\right| > C S_\varepsilon(l).$$

Let us denote

$$D_l(s) = \mathbf{P}_f\left[\left|\hat{f}_{\kappa(s-1)\wedge l}(t) - \hat{f}_l(t)\right| > C S_\varepsilon(l)\right].$$

Using this notation it follows:

$$D(s) \leq \sum_{l \in \mathcal{Z}_\varepsilon, l \succeq \kappa(s-1)} D_l(s).$$

Now, we have to control $D_l(s)$. Set $l \succeq \kappa(s-1)$. We have, using lemmas (3), (4) and (5):

$$\begin{aligned} D_l(s) &\leq \mathbf{P}_f\big[|b_{\kappa(s-1)\wedge l}(t) - b_l(t)| \\ &\qquad +\sigma_\varepsilon(\kappa(s-1)\wedge l)|\xi(\kappa(s-1)\wedge l)| \\ &\qquad +\sigma_\varepsilon(l)|\xi(l)| > CS_\varepsilon(l)\big] \\ &\leq \mathbf{P}_f\big[2B^{\beta,L}(\kappa(s-1)) \\ &\qquad +\sigma_\varepsilon(\kappa(s-1)\wedge l)|\xi(\kappa(s-1)\wedge l)| \\ &\qquad +\sigma_\varepsilon(l)|\xi(l)| > CS_\varepsilon(l)\big] \\ &\leq \mathbf{P}_f\big[2C^*S_\varepsilon(\kappa(s-1)) \\ &\qquad +\sigma_\varepsilon(\kappa(s-1)\wedge l)|\xi(\kappa(s-1)\wedge l)| \\ &\qquad +\sigma_\varepsilon(l)|\xi(l)| > CS_\varepsilon(l)\big]. \end{aligned}$$

Using lemma (3), it follows

$$\begin{aligned} D_l(s) &\leq \mathbf{P}_f\left[|\xi(\kappa(s-1)\wedge l)| + |\xi(l)| > (C-2C^*)\sqrt{1+|l|\ln 2}\right] \\ &\leq 2\mathbf{P}\left[|\mathcal{N}(0,1)| > \frac{C-2C^*}{2}\sqrt{1+|l|\ln 2}\right] \\ &\leq 2^{-\tilde{C}|l|+1}, \end{aligned}$$

where $\tilde{C} = (C-2C^*)^2/8$. Thus,

$$D(s) \leq 2 \sum_{l\in\mathcal{Z}_\varepsilon, l\succeq\kappa(s-1)} 2^{-\tilde{C}|l|}. \tag{6}$$

Control of $\boldsymbol{R(s,p)}$. Let us recall that

$$R(s,p) = \mathbf{E}_f\left[\left|\hat{f}_{\hat{k}}(t) - f(t)\right|^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\right].$$

The main idea is to decompose

$$\left|\hat{f}_{\hat{k}}(t) - f(t)\right|$$

by introducing $\hat{f}_{\kappa(s)}(t)$. In order to do that, we have to introduce $\hat{f}_{\hat{k}\wedge\kappa(s)}(t)$. Let us write:

$$\begin{aligned} \left|\hat{f}_{\hat{k}}(t) - f(t)\right| &\leq \left|\hat{f}_{\hat{k}}(t) - \hat{f}_{\hat{k}\wedge\kappa(s)}(t)\right| \\ &\quad + \left|\hat{f}_{\hat{k}\wedge\kappa(s)}(t) - \hat{f}_{\kappa(s)}(t)\right| \\ &\quad + \left|\hat{f}_{\kappa(s)}(t) - f(t)\right|. \end{aligned}$$

It is easy to prove that, if $s \leq s_{\max}$ then $\kappa(s)$ belongs to $\mathcal{Z}_\varepsilon$. This is a very important point. We will consider only $s \leq s_{\max}$. Let us recall that, if $s > s_{\max}$ then $D(s) = 0$.

Let us denote:

$$\begin{cases} I_1 &= |\hat{f}_{\hat{k}}(t) - \hat{f}_{\hat{k}\wedge\kappa(s)}(t)|\mathbf{1}_{\{\hat{k}\in B_1(s)\}} \\ I_2 &= |\hat{f}_{\hat{k}\wedge\kappa(s)}(t) - \hat{f}_{\kappa(s)}(t)|\mathbf{1}_{\{\hat{k}\in B_1(s)\}} \\ I_3 &= |\hat{f}_{\kappa(s)}(t) - f(t)|\mathbf{1}_{\{\hat{k}\in B_1(s)\}}. \end{cases}$$

We have:

$$R(s,p) \leq \left(3^{p-1} \vee 1\right) \left(\mathbf{E}_f[I_1^p] + \mathbf{E}_f[I_2^p] + \mathbf{E}_f[I_3^p]\right).$$

a) Let us control $\mathbf{E}_f[I_3^p]$. Using lemmas (3), (4) and (5), we have:

$$\begin{aligned} \mathbf{E}_f[I_3^p] &= \mathbf{E}_f\left[|\hat{f}_{\kappa(s)}(t) - f(t)|\mathbf{1}_{\{\hat{k}\in B_1(s)\}}\right] \\ &\leq \mathbf{E}_f\left[\left(|b_{\kappa(s)}(t)| + \sigma_\varepsilon(\kappa(s))|\xi(\kappa(s))|\right)^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\right] \\ &\leq \mathbf{E}_f\left[\left(B^{\beta,L}(\kappa(s)) + \sigma_\varepsilon(\kappa(s))|\xi(\kappa(s))|\right)^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\right] \\ &\leq \mathbf{E}_f\left[(C^* S_\varepsilon(\kappa(s)) + \sigma_\varepsilon(\kappa(s))|\xi(\kappa(s))|)^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\right]. \end{aligned}$$

Thus, we obtain:

$$\begin{aligned} \mathbf{E}_f[I_3^p] \leq\ & \left(2^{p-1} \vee 1\right) (C^*)^p S_\varepsilon^p(\kappa(s)) \\ & + \left(2^{p-1} \vee 1\right) \mathbf{E}_f\left[(\sigma_\varepsilon(\kappa(s))|\xi(\kappa(s))|)^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\right]. \end{aligned} \tag{7}$$

b) Let us control $\mathbf{E}_f[I_2^p]$. First, let us remark that:

- $\hat{k}$ belongs to $\mathcal{A}$. By definition of $\hat{k}$.
- $|\kappa(s)| \geq |\hat{k}|$. Because $\hat{k}$ belongs to $B_1(s)$.
- $\kappa(s)$ belongs to $\mathcal{Z}_\varepsilon$. Thanks to lemma **??**.

Thus, the construction of our procedure implies that

$$\mathbf{E}_f[I_2^p] \leq C^p S_\varepsilon^p(\kappa(s)). \tag{8}$$

c) Let us control $\mathbf{E}_f[I_1^p]$. Using lemmas 3, 4 and 5, it is easy to see that:

$$\begin{aligned} I_1 \leq\ & 2C^* S_\varepsilon(\kappa(s))\mathbf{1}_{\{\hat{k}\in B_1(s)\}} \\ & + \sigma_\varepsilon(\hat{k})|\xi(\hat{k})|\mathbf{1}_{\{\hat{k}\in B_1(s)\}} \\ & + \sigma_\varepsilon(\hat{k}\wedge\kappa(s))|\xi(\hat{k}\wedge\kappa(s))|\mathbf{1}_{\{\hat{k}\in B_1(s)\}}. \end{aligned}$$

Thus, we obtain:

$$\begin{aligned}\mathbf{E}_f[I_1^p] \;\leq\; & \left(3^{p-1}\vee 1\right)(2C^*)^p S_\varepsilon^p(\kappa(s)) \\ & + \left(3^{p-1}\vee 1\right)\mathbf{E}_f\left[\left(\sigma_\varepsilon(\hat{k})|\xi(\hat{k})|\right)^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\right] \\ & + \left(3^{p-1}\vee 1\right)\mathbf{E}_f\left[\left(\sigma_\varepsilon(\hat{k}\wedge\kappa(s))|\xi(\hat{k}\wedge\kappa(s))|\right)^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\right] \quad (9)\end{aligned}$$

Using inequalities (7)–(8)–(9), we obtain:

$$\begin{aligned}R(s,p) \;\leq\; & \left(\left(2^{p-1}\vee 1\right)(C^*)^p + C^p + \left(3^{p-1}\vee 1\right)(2C^*)^p\right) S_\varepsilon^p(\kappa(s)) \\ & + \left(2^{p-1}\vee 1\right)\mathbf{E}_f\left[(\sigma_\varepsilon(\kappa(s))|\xi(\kappa(s))|)^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\right] \\ & + \left(3^{p-1}\vee 1\right)\mathbf{E}_f\left[\left(\sigma_\varepsilon(\hat{k})|\xi(\hat{k})|\right)^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\right] \\ & + \left(3^{p-1}\vee 1\right)\mathbf{E}_f\left[\left(\sigma_\varepsilon(\hat{k}\wedge\kappa(s))|\xi(\hat{k}\wedge\kappa(s))|\right)^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\right] \quad (10)\end{aligned}$$

Thus, we have to control the expectations in the last inequality.
It is easy to control the first one:

$$\mathbf{E}_f\left[(\sigma_\varepsilon(\kappa(s))|\xi(\kappa(s))|)^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\right] \leq \sigma_\varepsilon(\kappa(s))\mathbf{E}\left[|\mathcal{N}(0,1)|^p\right].$$

Now, let us explain how to control the others. Let us denote $\tilde{k} = \hat{k}\wedge\kappa(s)$ and

$$\Lambda_k = \left\{|\xi(k\wedge\kappa(s))| > 2\sqrt{1+|k|\ln 2}\right\}.$$

Now, let us calculate:

$$\begin{aligned}(*) \;=\; & \mathbf{E}_f\left[\left(\sigma_\varepsilon(\hat{k}\wedge\kappa(s))|\xi(\hat{k}\wedge\kappa(s))|\right)^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\right] \\ =\; & \mathbf{E}_f\left[\left(\sigma_\varepsilon(\tilde{k})|\xi(\tilde{k})|\right)^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\left(\mathbf{1}_{\{\Lambda_{\hat{k}}\}} + \mathbf{1}_{\{\Lambda_{\hat{k}}^c\}}\right)\right] \\ \leq\; & \mathbf{E}_f\left[\left(\sigma_\varepsilon(\tilde{k})|\xi(\tilde{k})|\right)^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\mathbf{1}_{\{\Lambda_{\hat{k}}\}}\right] \\ & + \mathbf{E}_f\left[\left(2\sigma_\varepsilon(\tilde{k})\sqrt{1+|\hat{k}|\ln 2}\right)^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\right] \\ \leq\; & \sigma_\varepsilon^p(\kappa(s))\mathbf{E}_f\left[|\xi(\tilde{k})|^p \mathbf{1}_{\{\Lambda_{\hat{k}}\}}\right] + (2S_\varepsilon(\kappa(s)))^p.\end{aligned}$$

Let us denote $m_l = \mathbf{E}\left[|\mathcal{N}(0,1)|^l\right]$. We obtain:

$$(*) \leq \sigma_\varepsilon^p(\kappa(s)) m_{2p}^{1/2}\mathbf{P}_f[\Lambda_{\hat{k}}] + (2S_\varepsilon(\kappa(s)))^p.$$

Moreover we have:

$$\begin{aligned}
\mathbf{P}_f[\Lambda_{\hat{k}}] &\leq \mathbf{P}_f\left[\bigcup_{k\in\mathcal{Z}_\varepsilon} \Lambda_k\right] \\
&\leq \sum_{k\in\mathcal{Z}_\varepsilon} 2^{-|k|} \\
&\leq \sum_{n=0}^{\infty}\sum_{k\in\mathcal{Z}_n} 2^{-n} \\
&\leq \sum_{n=0}^{\infty}(\#\mathcal{Z}_n)2^{-n} < +\infty.
\end{aligned}$$

Let us denote $|\mathcal{Z}| = \sum_{n=0}^{\infty}(\#\mathcal{Z}(n))2^{-n}$. We obtain:

$$(*) \leq \sigma_\varepsilon^p(\kappa(s))m_{2p}^{1/2}|\mathcal{Z}| + (2S_\varepsilon(\kappa(s)))^p \leq \left(2^p + m_{2p}^{1/2}|\mathcal{Z}|\right) S_\varepsilon^p(\kappa(s)).$$

It is not difficult to obtain a similar result for the last expectation:

$$\mathbf{E}_f\left[\left(\sigma_\varepsilon(\hat{k})|\xi(\hat{k})|\right)^p \mathbf{1}_{\{\hat{k}\in B_1(s)\}}\right] \leq \left(2^p + m_{2p}^{1/2}|\mathcal{Z}|\right) S_\varepsilon^p(\kappa(s)).$$

Finaly, using (10) and the control of the expaectations we obtain:

$$R(s,p) \leq C_p S_\varepsilon^p(\kappa(s)) \tag{11}$$

where C_p is a constant depending only on p and $|\mathcal{Z}|$.
Back to our problem. Now, we can conclude. Let us recall that:

$$(**) = \mathbf{E}_f\left[\left(\left|f_\varepsilon^\Phi(t) - f(t)\right|\right)^q\right] \leq R(0,q) + \sum_{s\geq 1}\sqrt{R(s,2q)D(s)}.$$

Thus, we obtain — using (6) and (11):

$$(**) \quad \leq \quad C_q S_\varepsilon^q(\kappa) + (2C_{2q})^{1/2}\sum_{s\geq 1}\sqrt{S_\varepsilon^{2q}(\kappa(s))\sum_{l\in\mathcal{Z}_\varepsilon, l\succeq\kappa(s-1)} 2^{-\tilde{C}|l|}}.$$

Let us recall that $\tilde{C} = 3q+2$ and that:

$$S_\varepsilon(\kappa(s)) = S_\varepsilon(0)2^{|\kappa(s)|}\sqrt{1+|\kappa(s)|\ln 2}.$$

Thus:

$$\begin{aligned}
S_\varepsilon^{2q}(\kappa(s))2^{-3q|\kappa(s-1)|} &\leq S_\varepsilon^{2q}(0)2^{2q|\kappa(s)|}(1+|\kappa(s)|\ln 2)^q 2^{-3q|\kappa(s)|} \\
&\leq S_\varepsilon^{2q}(0)2^{2q(|\kappa(s)|-|\kappa(s-1)|)}\left(\frac{1+|\kappa(s)|\ln 2}{2^{|\kappa(s-1)|}}\right)^q \\
&\leq 3^{dq}S_\varepsilon^{2q}(0).
\end{aligned}$$

Now, it is easy to see that (we do not recall that $l \in \mathcal{Z}_\varepsilon$):

$$\begin{aligned}
(**) \quad &\leq \quad C_q S_\varepsilon^q(\kappa) + (3^{dq} 2 C_{2q})^{1/2} S_\varepsilon^q(0) \sum_{s \geq 1} \sqrt{\sum_{l \succeq \kappa(s-1)} 2^{-3q(|l|-|\kappa(s-1))-2|l|}} \\
&\leq \quad C_q S_\varepsilon^q(\kappa) + (3^{dq} 2 C_{2q})^{1/2} S_\varepsilon^q(0) \sum_{s \geq 1} \sqrt{\sum_{l \succeq \kappa(s-1)} 2^{-2|l|}} \\
&\leq \quad C_q S_\varepsilon^q(\kappa) + (3^{dq} 2 C_{2q})^{1/2} S_\varepsilon^q(0) \sum_{s \geq 1} \sum_{l \succeq \kappa(s-1)} 2^{-|l|}
\end{aligned}$$

Now, we have to prove that:

$$\sum_{s \geq 1} \sum_{l \succeq \kappa(s-1)} 2^{-|l|} < +\infty.$$

Let us calculate:

$$\begin{aligned}
\sum_{s \geq 1} \sum_{l \succeq \kappa(s-1)} 2^{-|l|} &= \sum_{s \geq 0} \sum_{n \geq |\kappa(s)|} (\# \mathcal{Z}(n))\, 2^{-n} \\
&= \sum_{s \geq 0} \sum_{n \geq s} (\# \mathcal{Z}(n))\, 2^{-n} \\
&= \sum_{n \geq 0} \sum_{s \leq n} (\# \mathcal{Z}(n))\, 2^{-n} \\
&= \sum_{n \geq 0} \frac{n(n+1)}{2} (\# \mathcal{Z}(n))\, 2^{-n} < +\infty.
\end{aligned}$$

Let us denote $\|\mathcal{Z}\|$ this constant. Finaly, if we remeber that $\kappa = k(\beta, L, \varepsilon)$ and that $S_\varepsilon(0) \leq S_\varepsilon(\kappa)$, we obtain the following result:

$$\mathbf{E}_f \left[\left(\left| f_\varepsilon^\Phi(t) - f(t) \right| \right)^q \right] \leq \left(C_q + (3^{dq} 2 C_{2q})^{1/2} \|\mathcal{Z}\| \right) S_\varepsilon(k(\beta, L, \varepsilon)).$$

As $S_\varepsilon(k(\beta, L, \varepsilon)) = \varphi_\varepsilon(\beta, L)$, result follows.

B) $\mathcal{A}$ is empty

This case is simpler. We have to control:

$$\mathbf{E}_f \left[|f(t)|^q \mathbf{1}_{\{\mathcal{A} = \emptyset\}} \right] \leq L^q \mathbf{P}_f[\mathcal{A} = \emptyset].$$

Moreover, we can assume that there exist $s \in \mathbf{N}$ such that $\kappa(s) = N_\varepsilon$ and $\kappa(s) \in \mathcal{Z}_\varepsilon$. The fact that $\mathcal{A}$ is empty implies that $\kappa(s)$ is not in $\mathcal{A}$, thus:

$$\mathbf{P}_f[\mathcal{A} = \emptyset] \leq \mathbf{P}_f[\kappa(s) \notin \mathcal{A}].$$

The same quantity was controlled by formula (6). Then, it is easy to obtain that:

$$\mathbf{P}_f[\mathcal{A} = \emptyset] \leq \left(\sum_{n \geq 0} (\# \mathcal{Z}_{2n}) \, 2^{-\tilde{C} n + 1} \right) 2^{-\tilde{C}|\kappa(s)|}.$$

The last thing we have to observe is that $2^{-\tilde{C}|\kappa(s)|} \leq S_\varepsilon(\kappa)$.

7 Proof of theorem 2

Let us consider another admissible family $\Psi = \{\psi_\varepsilon(\beta, L)\}_{(\beta,L)\in\mathcal{B}\times\mathcal{I}}$ and f_ε^Ψ an estimator satisfying (A.U.B) with Ψ.

To prove that Φ is the adaptive rate, it is enough to prove the following assertion:

Lemma 7. *Set $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathcal{B}$ and $\beta = (\beta_1, \ldots, \beta_d) \in \mathcal{B}$ such that $\bar{\alpha} < \bar{\beta}$. Set L_α and L_β in $\mathcal{I}$. If*

$$\frac{\psi_\varepsilon(\alpha, L_\alpha)}{\varphi_\varepsilon(\alpha, L_\alpha)} \xrightarrow[\varepsilon \to 0]{} 0,$$

then,

$$\frac{\psi_\varepsilon(\beta, L_\beta)}{\varphi_\varepsilon(\beta, L_\beta)} \times \frac{\psi_\varepsilon(\alpha, L_\alpha)}{\varphi_\varepsilon(\alpha, L_\alpha)} \xrightarrow[\varepsilon \to 0]{} +\infty.$$

Indeed, let us remark, first, that we cannot improve $\varphi_\varepsilon(b, L)$ because it is the minimax rate of convergence on $H(b, L)$ for all L.

Next, let us suppose that there exists (β_0, L_0) such that $\psi_\varepsilon(\beta_0, L_0)$ improves $\varphi_\varepsilon(\beta_0, L_0)$. Using the previous lemma, it is easy to sea that $\mathcal{I}_0(\Psi/\Phi) \subset \mathcal{B}(\bar{\beta}_0) \times \mathcal{I}$. Indeed, let us suppose that there exists $(\beta_1, L1)$ such that $\psi_\varepsilon(\beta_1, L_1)$ improves $\varphi_\varepsilon(\beta_1, L_1)$ and $\bar{\beta}_1 < \bar{\beta}_0$ then we obtain:

$$\frac{\psi_\varepsilon(\beta_1, L_{\beta_1})}{\varphi_\varepsilon(\beta_1, L_{\beta_1})} \times \frac{\psi_\varepsilon(\beta_0, L_{\beta_0})}{\varphi_\varepsilon(\beta_0, L_{\beta_0})} \xrightarrow[\varepsilon \to 0]{} +\infty.$$

In particular $\psi_\varepsilon(\beta_0, L_{\beta_0})/\varphi_\varepsilon(\beta_0, L_{\beta_0})$ tends to $+\infty$ which it is impossible.

On the other hand, it is easy to sea that $\bigcup_{\gamma > \beta_0} \mathcal{B}(\gamma) \times \mathcal{I} \subset \mathcal{I}_\infty(\Psi/\Phi)$.

Lemma 7 is a corrolary of the following proposition:

Proposition 1. *Set $(\alpha, \beta) \in \mathcal{B}^2$ such that $\bar{\alpha} < \bar{\beta}$ and $(L_\alpha, L_\beta) \in \mathcal{I}^2$. Let us define, for any estimator $\tilde{f}_\varepsilon(\cdot)$ which satisfies (A.U.B.) with Ψ and for all $\nu < 2(\bar{\beta} - \bar{\alpha})/((2\bar{\alpha} + 1)(2\bar{\beta} + 1))$ the following quantiy:*

$$\begin{aligned} R_\varepsilon^{(q)}(\tilde{f}_\varepsilon) &= \sup_{f \in H(\alpha, L_\alpha)} \mathbf{E}_f \left[\left(\varphi_\varepsilon^{-1}(\alpha, L_\alpha) |\tilde{f}_\varepsilon(t) - f(t)| \right)^q \right] \\ &\quad + \sup_{f \in H(\beta, L_\beta)} \mathbf{E}_f \left[\left(\varepsilon^\nu \varphi_\varepsilon^{-1}(\beta, L_\beta) |\tilde{f}_\varepsilon(t) - f(t)| \right)^q \right]. \end{aligned}$$

If $\psi_\varepsilon(\alpha, L_\alpha)/\varphi_\varepsilon(\alpha, L_\alpha)$ tends to 0 as $\varepsilon \to 0$, then, we have:

$$\liminf_{\varepsilon\to 0} R_\varepsilon^{(q)}(\tilde{f}_\varepsilon) > 0.$$

Proof. As $\psi_\varepsilon(\delta, L_\delta)$ and $\varphi_\varepsilon(\delta, L_\delta)$ do not depend, in order, on L_δ ($\delta \in \{\alpha, \beta\}$), we will denote, for simplicity:

$$\psi_\varepsilon(\delta) \triangleq \psi_\varepsilon(\delta, L_\delta) \text{ and } \varphi_\varepsilon(\delta) \triangleq \left(\varepsilon\sqrt{\ln 1/\varepsilon}\right)^{\frac{2\bar{\delta}}{2\bar{\delta}+1}}.$$

Set $\varkappa$ a positive parameter to be chosen. We consider $h_i = h_i(\varepsilon)$ defined by the formula

$$h_i = \left(\varkappa\varepsilon\sqrt{\ln \varepsilon^{-1}}\right)^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}\frac{1}{\alpha_i}},$$

and we consider the two following functions:

$$\begin{cases} f_0 &= 0 \\ f_1(x) &= L_\alpha \varkappa^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}} \varphi_\varepsilon(\alpha) f\left(\frac{x_1-t_1}{h_1}, \dots, \frac{x_d-t_d}{h_d}\right) \end{cases}$$

where f belongs to $H(\alpha, 1)$. Hence f_1 belongs to $H(\alpha, L_\alpha)$ and, if $\mathbf{E}_0$ and $\mathbf{E}_1$ denote respectively $\mathbf{E}_{f_0}$ and $\mathbf{E}_{f_1}$, we have:

$$\begin{aligned} R_\varepsilon^{(q)}(\tilde{f}_\varepsilon) &\geq \mathbf{E}_0\left|\varepsilon^\nu\varphi_\varepsilon^{-1}(\beta)\tilde{f}_\varepsilon(t)\right|^q + \mathbf{E}_1\left|\varphi_\varepsilon^{-1}(\alpha)(\tilde{f}_\varepsilon(t) - f_1(t))\right|^q \\ &\geq \mathbf{E}_0\left|\varepsilon^\nu\varphi_\varepsilon^{-1}(\beta)\tilde{f}_\varepsilon(t)\right|^q + \mathbf{E}_1\left|\varphi_\varepsilon^{-1}(\alpha)\tilde{f}_\varepsilon(t) - z\right|^q, \end{aligned}$$

where z denote $L\varkappa^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}} f(0)$. For simplicity, we consider the following notations:

$$\lambda_\varepsilon = \varepsilon^\nu \frac{\varphi_\varepsilon(\alpha)}{\varphi_\varepsilon(\beta)} = \varepsilon^\nu\left(\varepsilon\sqrt{\ln\frac{1}{\varepsilon}}\right)^{-\varrho} \text{ where } \varrho = \frac{2(\bar{\beta}-\bar{\alpha})}{(2\bar{\beta}+1)(2\bar{\alpha}+1)}.$$

and

$$\tilde{\theta} = \varphi_\varepsilon^{-1}(\alpha)|\tilde{f}_\varepsilon(t)|.$$

Thus, we have:

$$R_\varepsilon^{(q)}(\tilde{f}_\varepsilon) \geq \mathbf{E}_0\left|\lambda_\varepsilon\tilde{\theta}\right|^q + \mathbf{E}_1\left|\tilde{\theta} - z\right|^q$$

By changing the probability measure, we obtain:

$$R_\varepsilon^{(q)}(\tilde{f}_\varepsilon) \geq \mathbf{E}_1\left[\left|\lambda_\varepsilon\tilde{\theta}\right|^q Z_\varepsilon + \left|\tilde{\theta} - z\right|^q\right]$$

where Z_ε denotes the classical likelyhood ratio:

$$Z_\varepsilon = \frac{d\mathbf{P}_0}{d\mathbf{P}_1}(\mathcal{X}^{(\varepsilon)}).$$

Now, consider the following event for $\delta > 0$:

$$\Lambda = \left\{ |\tilde{\theta}| > \delta \right\}$$

Clearly, if δ is small enough:

$$R_\varepsilon^{(q)}(\tilde{f}_\varepsilon) \geq \mathbf{E}_1 \left[(\delta\lambda_\varepsilon)^q Z_\varepsilon \mathbf{1}_{\{\Lambda\}} + (z-\delta)^q \mathbf{1}_{\{\Lambda^c\}} \right].$$

But,

$$\begin{aligned} Z_\varepsilon &= \exp\left(-\frac{1}{\varepsilon} \int_{\mathbf{R}^d} f_1(u) dW(u) - \frac{1}{2\varepsilon^2} \|f_1\|^2 \right) \\ &= \exp\left(-\frac{\|f_1\|}{\varepsilon} \xi - \frac{1}{2} \left(\frac{\|f_1\|}{\varepsilon} \right)^2 \right), \end{aligned}$$

where $\xi \sim \mathcal{N}(0,1)$. Hence, if the event

$$\Lambda^a = \{ |\xi| \leq a \}$$

occurs, we can deduce that:

$$\begin{aligned} Z_\varepsilon &\geq \exp\left(-\frac{\|f_1\|}{\varepsilon} a - \frac{1}{2} \frac{\|f_1\|^2}{\varepsilon^2} \right) \\ &\geq \exp\left(-\frac{1}{2} \left(\frac{\|f_1\|}{\varepsilon} + a \right)^2 \right) \end{aligned}$$

Then, we have:

$$\begin{aligned} R_\varepsilon^{(q)}(\tilde{f}_\varepsilon) &\geq \mathbf{E}_1 \left[(\delta\lambda_\varepsilon)^q Z_\varepsilon \mathbf{1}_{\{\Lambda \cap \Lambda^a\}} + (z-\delta)^q \mathbf{1}_{\{\Lambda^c\}} \right] \\ &\geq \mathbf{E}_1 \left[(\delta\lambda_\varepsilon)^q \exp\left(-\frac{1}{2} \left(\frac{\|f_1\|}{\varepsilon} + a \right)^2 \right) \mathbf{1}_{\{\Lambda \cap \Lambda^a\}} + (z-\delta)^q \mathbf{1}_{\{\Lambda^c\}} \right]. \end{aligned}$$

Let us remark that:

$$\frac{\|f_1\|}{\varepsilon} = L\varkappa \|f\| \sqrt{\ln \frac{1}{\varepsilon}}.$$

If we chose

$$a = L\varkappa \|f\| \sqrt{\ln \frac{1}{\varepsilon}} \wedge 1,$$

then we obtain:

$$\begin{aligned}
R_\varepsilon^{(q)}(\tilde{f}_\varepsilon) &\geq \mathbf{E}_1\left[(\delta\lambda_\varepsilon)^q \exp\left(-\left(\frac{\|f_1\|}{\varepsilon}\right)^2\right)\mathbf{1}_{\{\Lambda\cap\Lambda^a\}} + (z-\delta)^q\mathbf{1}_{\{\Lambda^c\}}\right] \\
&\geq \mathbf{E}_1\left[(\delta\lambda_\varepsilon)^q\varepsilon^{(L_\alpha\varkappa\|f\|)^2}\mathbf{1}_{\{\Lambda\cap\Lambda^a\}} + (z-\delta)^q\mathbf{1}_{\{\Lambda^c\}}\right] \\
&\geq \mathbf{E}_1\left[(\delta\eta_\varepsilon)^q\varepsilon^{q(\nu-\varrho)+(L_\alpha\varkappa\|f\|)^2}\mathbf{1}_{\{\Lambda\cap\Lambda^a\}} + (z-\delta)^q\mathbf{1}_{\{\Lambda^c\}}\right],
\end{aligned}$$

where $\eta_\varepsilon = \left(\ln\frac{1}{\varepsilon}\right)^{-\varrho/2}$.

Let us introduce

$$t_\varepsilon = \frac{q}{\ln\frac{1}{\varepsilon}}\left(\ln\frac{1}{\delta\eta_\varepsilon} + \ln AL_\alpha f(0)\right) \to 0,$$

where

$$A = \left(\frac{\sqrt{q(\varrho-\nu)}}{L_\alpha\|f\|}\right)^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}},$$

and let us chose:

$$\varkappa = \frac{\sqrt{q(\varrho-\nu)-t_\varepsilon}}{L_\alpha\|f\|}.$$

Using this choice of $\varkappa$, we obtain that:

$$(L_\alpha\varkappa\|f\|)^2 = q\varrho - t_\varepsilon \text{ and } (\delta\eta_\varepsilon)^q\varepsilon^{-q\varrho+(L_\alpha\varkappa\|f\|)} = (AL_\alpha f(0))^q$$

Thus, we obtain:

$$\begin{aligned}
R_\varepsilon^{(q)}(\tilde{f}_\varepsilon) &\geq \mathbf{E}_1\left[(AL_\alpha f(0))^q\mathbf{1}_{\{\Lambda\cap\Lambda^a\}} + (\varkappa^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}}L_\alpha f(0)-\delta)^q\mathbf{1}_{\{\Lambda^c\}}\right] \\
&> \mathbf{E}_1\left[(\varkappa^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}}L_\alpha f(0)-\delta)^q\left(\mathbf{1}_{\{\Lambda+\Lambda^c\}}\right)\mathbf{1}_{\{\Lambda^a\}}\right] \\
&\geq (\varkappa^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}}L_\alpha f(0)-\delta)^q\mathbf{P}_1\left[\Lambda^a\right] \\
&\geq (\varkappa^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}}L_\alpha f(0)-\delta)^q\mathbf{P}\left[|\xi|\geq 1\right].
\end{aligned}$$

And, then:

$$\liminf_{\varepsilon\to 0} R_\varepsilon^{(q)}(\tilde{f}_\varepsilon) \geq \left(L_\alpha^{\frac{1}{2\bar{\alpha}+1}}\frac{f(0)}{\|f\|^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}}}\left(q(\varrho-\nu)\right)^{\frac{\bar{\alpha}}{2\bar{\alpha}+1}}\right)^q\mathbf{P}\left[|\xi|\geq 1\right] > 0.$$

Proposition is proved.

A Proof of lemma 1

Let us denote $\mathcal{H} = (\mathbf{R}_+^*)^d$ and $\partial\mathcal{H} = \{h \in (\mathbf{R}_+^*)^d; \exists i \in [\![1;d]\!] h_i = 0\} \cup \{\infty\}$.

Let us recall that $\varphi_\varepsilon^{\beta,L} = b^{\beta,L} + s_\varepsilon^{\beta,L}$. It is easy to prove the following assertion:

$$\varphi_\varepsilon^{\beta,L}(h) \xrightarrow[h\to\partial\mathcal{H}]{} +\infty.$$

Thus, it is enough, to prove Lemma, to prove that $h(\beta, L, \varepsilon)$ is the unique point of $\mathcal{H}$ such that $\nabla\varphi_\varepsilon^{\beta,L} = 0$. Let us fix $i \in [\![1;d]\!]$, and let us calculate:

$$\partial_i\varphi_\varepsilon^{\beta,L}(h) = L\beta_i\lambda_i(\beta)h_i^{\beta_i-1} - \frac{\|K\|\varepsilon}{2\left(\prod_{j=1}^d h_j\right)^{\frac{1}{2}}}\frac{1}{h_i}\rho_\varepsilon(\beta, L).$$

To simplify notations, as β and L are fixed, we will denote λ_i instead of $\lambda_i(\beta)$. It follows that:

$$\partial_i\varphi_\varepsilon^{\beta,L}(h) = 0 \Leftrightarrow h_i^{\beta_i} = (\lambda_i\beta_i)^{-1}\frac{\|K\|}{2L}\frac{\varepsilon\rho_\varepsilon(\beta, L)}{\left(\prod_{j=1}^d h_j\right)^{\frac{1}{2}}}.$$

It is easy to deduce, from the previous equality, the following expression for h_i:

$$h_i = (\lambda_i\beta_i)^{\frac{-1}{\beta_i}}\left(\frac{\|K\|}{2L}\frac{\varepsilon\rho_\varepsilon(\beta, L)}{\left(\prod_{j=1}^d h_j\right)^{\frac{1}{2}}}\right)^{\frac{1}{\beta_i}}.$$

A simple computation prove that:

$$\left(\prod_{i=1}^d h_i\right)^{\frac{1}{2}} = \left(\prod_{i=1}^d (\lambda_i\beta_i)^{\frac{-1}{\beta_i}}\right)^{\frac{\bar\beta}{2\bar\beta+1}}\left(\frac{\|K\|}{2L}\varepsilon\rho_\varepsilon(\beta, L)\right)^{\frac{1}{2\bar\beta+1}}.$$

The follwing equality follows easily from previous equalities:

$$h_i = \gamma_i\left(\frac{\|K\|\Gamma}{2L}\varepsilon\rho_\varepsilon(\beta, L)\right)^{\frac{2\bar\beta}{2\bar\beta+1}\frac{1}{\beta_i}}$$

where

$$\left\{\begin{array}{lcl}\gamma_i & = & (\lambda_i\beta_i)^{\frac{-1}{\beta_i}}\\ \Gamma & = & \left(\prod_{i=1}^d \gamma_i\right)^{\frac{1}{2}}.\end{array}\right.$$

Conversely, it is easy to proved that $h \in \mathcal{H}$ given by the previous formulas is such that $\nabla\varphi_\varepsilon^{\beta,L}(h) = 0$. This result implies that the problem is solved. Lemma is proved.

B Proof of lemma 2

This lemma is the most technical one. It will be proved in two steps.

Step 1. *Set* $(\beta, L) \in \mathcal{B} \times \mathcal{I}$. *We have:*

$$0 \leq \sum_{i=1}^{d} (k_i(\beta, L, \varepsilon) + 1) \leq N_\varepsilon$$

where N_ε *is defined by:*

$$\left\lfloor 2 \left(\frac{2\bar{b}}{2\bar{b}+1} \ln \frac{l_*}{\|K\|\varepsilon} + \ln \frac{l^*}{l_*} \right) \right\rfloor + 1.$$

Proof. Let us denote

$$\begin{aligned} x_\varepsilon(\beta, L) &= \frac{4(\bar{b}-\bar{\beta})}{(2\bar{b}+1)(2\bar{\beta}+1)} \ln \frac{L}{\|K\|\varepsilon} + \frac{2}{2\bar{b}+1} \ln \frac{L}{l_*} \\ &= \frac{4(\bar{b}-\bar{\beta})}{(2\bar{b}+1)(2\bar{\beta}+1)} \ln \frac{l_*}{\|K\|\varepsilon} + \frac{2}{2\bar{\beta}+1} \ln \frac{L}{l_*}. \end{aligned}$$

Using this notation, it is easy to proof that

$$\ln \prod_{i=1}^{d} \frac{h_i^*(\varepsilon)}{h_i(\beta, L, \varepsilon)} = x_\varepsilon(\beta, L) - \frac{1}{2\bar{\beta}+1} \ln(1 + x_\varepsilon(\beta, L)).$$

Moreover,

$$\frac{1}{2\bar{\beta}+1} \ln(1 + x_\varepsilon(\beta, L)) \leq \ln(1 + x_\varepsilon(\beta, L)) \leq x_\varepsilon(\beta, L),$$

thus

$$\ln \prod_{i=1}^{d} \frac{h_i^*(\varepsilon)}{h_i(\beta, L, \varepsilon)} \geq 0.$$

This result implies that:

$$\sum_{i=1}^{d} (k_i(\beta, L, \varepsilon) + 1) \geq 0.$$

On the other hand,

$$\begin{aligned} \ln \prod_{i=1}^{d} \frac{h_i^*(\varepsilon)}{h_i(\beta, L, \varepsilon)} &\leq x_\varepsilon(\beta, L) \\ &= \frac{4(\bar{b}-\bar{\beta})}{(2\bar{b}+1)(2\bar{\beta}+1)} \ln \frac{l_*}{\|K\|\varepsilon} + \frac{2}{2\bar{\beta}+1} \ln \frac{L}{l_*} \\ &\leq \frac{4\bar{b}}{2\bar{b}+1} \ln \frac{l_*}{\|K\|\varepsilon} + 2 \ln \frac{l^*}{l_*} \end{aligned}$$

Step 2. *Set $(\beta, L) \in \mathcal{B} \times \mathcal{I}$ and let us denote $n = \sum_i (k_i(\beta, L, \varepsilon) + 1)$. Then, for all $i \in [\![1; d]\!]$, we have:*

$$|k_i(\beta, L, \varepsilon)| \leq \left(\frac{2\bar{b}+1}{2\bar{b}} \times \frac{\ln(1+\delta) + \sqrt{2\ln(1+\delta)}}{\ln(1+\delta)} \right) n + 1.$$

Proof. In this case, we can write

$$\begin{aligned} n\ln(2) &\geq \ln \prod_{i=1}^{d} \frac{h_i^*(\varepsilon)}{h_i(\beta, L, \varepsilon)} \\ &= x_\varepsilon(\beta, L) - \frac{1}{2\bar{\beta}+1} \ln(1 + x_\varepsilon(\beta, L)). \end{aligned}$$

Thus, we have:

$$x_\varepsilon(\beta, L) \leq n\ln(2) + \ln(1 + x_\varepsilon(\beta, L)).$$

Now, let us remark that, if x is such that $x \leq A + \ln(1+x)$ for a given constant $A > 0$, then $x \leq A + \sqrt{2A}$. Thus, we have

$$\begin{aligned} x_\varepsilon(\beta, L) &\leq n\ln(2) + \sqrt{2n\ln(2)} \\ &\leq n\left(\ln(2) + \sqrt{2\ln(2)}\right). \end{aligned}$$

Now, let us write

$$\begin{aligned} \ln \frac{h_i^*(\varepsilon)}{h_i(\beta, L, \varepsilon)} &= \left(\frac{2\bar{\beta}}{2\bar{\beta}+1} \frac{1}{\beta_i} - \frac{2\bar{b}}{2\bar{b}+1} \frac{1}{b_i} \right) \ln \frac{l_*}{\|K\|\varepsilon} \\ &\quad + \frac{2\bar{\beta}}{2\bar{\beta}+1} \frac{1}{\beta_i} \ln \frac{L}{l_*} \\ &\quad - \frac{\bar{\beta}}{2\bar{\beta}+1} \frac{1}{\beta_i} \ln(1 + x_\varepsilon(\beta, L)). \end{aligned}$$

Let us estimate this quatity.

Upper bound. First, using the fact that $\bar{\beta} \leq \beta_i$ for all i, we obtain:

$$\ln \frac{h_i^*(\varepsilon)}{h_i(\beta, L, \varepsilon)} \leq \left(\frac{2\bar{\beta}}{2\bar{\beta}+1} \frac{1}{\beta_i} - \frac{2\bar{b}}{2\bar{b}+1} \frac{1}{b_i} \right) \ln \frac{l_*}{\|K\|\varepsilon} + \frac{2}{2\bar{\beta}+1} \ln \frac{L}{l_*}.$$

On the other hand, it is easy to prove that:

$$\frac{2\bar{\beta}}{2\bar{\beta}+1} \frac{1}{\beta_i} - \frac{2\bar{b}}{2\bar{b}+1} \frac{1}{b_i} \leq \frac{2\bar{b}+1}{2\bar{b}} \frac{4(\bar{b} - \bar{\beta})}{(2\bar{b}+1)(2\bar{\beta}+1)}.$$

Indeed, let us write:

$$\begin{aligned}
\frac{2\bar\beta}{2\bar\beta+1}\frac{1}{\beta_i}-\frac{2\bar b}{2\bar b+1}\frac{1}{b_i} &= \frac{2\bar\beta}{2\bar\beta+1}\left(\frac{1}{\beta_i}-\frac{1}{b_i}\right)-\frac{1}{b_i}\left(\frac{2\bar b}{2\bar b+1}-\frac{2\bar\beta}{2\bar\beta+1}\right)\\
&\le \frac{2\bar\beta}{2\bar\beta+1}\left(\frac{1}{\beta_i}-\frac{1}{b_i}\right)\\
&\le \frac{2\bar\beta}{2\bar\beta+1}\left(\frac{1}{\bar\beta}-\frac{1}{\bar b}\right)\\
&\le \frac{2\bar\beta}{2\bar\beta+1}\frac{\bar b-\bar\beta}{\bar b\bar\beta}\\
&= \frac{2\bar b+1}{2\bar b}\frac{4(\bar b-\bar\beta)}{(2\bar b+1)(2\bar\beta+1)}.
\end{aligned}$$

Finally, we obtain:

$$\ln\frac{h_i^*(\varepsilon)}{h_i(\beta,L,\varepsilon)}\le\frac{2\bar b+1}{2\bar b}x_\varepsilon(\beta,L),$$

and thus

$$\begin{aligned}
k_i(\beta,L,\varepsilon) &\le \frac{1}{\ln(2)}\ln\frac{h_i^*(\varepsilon)}{h_i(\beta,L,\varepsilon)}\\
&\le \frac{2\bar b+1}{2\bar b}\times\frac{x_\varepsilon(\beta,L)}{\ln(2)}\\
&\le \left(\frac{2\bar b+1}{2\bar b}\times\frac{\ln(2)+\sqrt{2\ln(2)}}{\ln(2)}\right)n.
\end{aligned}$$

Lower bound. First, let us supose that the following fact is proved:

$$\frac{2\bar\beta}{2\bar\beta+1}\frac{1}{\beta_i}-\frac{2\bar b}{2\bar b+1}\frac{1}{b_i}\ge-\frac{1}{2\bar b}\frac{4(\bar b-\bar\beta)}{(2\bar b+1)(2\bar\beta+1)}. \tag{12}$$

Then, we obtain

$$\ln\frac{h_i^*(\varepsilon)}{h_i(\beta,L,\varepsilon)}\ge-\frac{1}{2\bar b}\frac{4(\bar b-\bar\beta)}{(2\bar b+1)(2\bar\beta+1)}\ln\frac{l_*}{\|K\|\varepsilon}-\frac{\bar\beta}{2\bar\beta+1}\frac{1}{\beta_i}\ln(1+x_\varepsilon(\beta,L)).$$

Using the inequality $x_\varepsilon(\beta,L)\ge\ln l_*/\|K\|\varepsilon$, it follows

$$\ln\frac{h_i^*(\varepsilon)}{h_i(\beta,L,\varepsilon)}\ge-\frac{1}{2\bar b}\left(x_\varepsilon(\beta,L)+\frac{2\bar\beta}{2\bar\beta+1}\frac{\bar b}{\beta_i}\ln(1+x_\varepsilon(\beta,L))\right).$$

And, then

$$\ln \frac{h_i^*(\varepsilon)}{h_i(\beta, L, \varepsilon)} \geq -\frac{2\bar{b}+1}{2\bar{b}} x_\varepsilon(\beta, L).$$

Finally:

$$\begin{aligned} k_i(\beta, L, \varepsilon) + 1 &\geq \frac{1}{\ln(2)} \ln \frac{h_i^*(\varepsilon)}{h_i(\beta, L, \varepsilon)} \\ &\geq -\left(\frac{2\bar{b}+1}{2\bar{b}} \times \frac{\ln(2) + \sqrt{2\ln(2)}}{\ln(2)} \right) n. \end{aligned}$$

To end the proof of this lema, we have to prove inequality (12) i.e.

$$(*) = \left(\frac{2\bar{\beta}}{2\bar{\beta}+1} \frac{1}{\beta_i} - \frac{2\bar{b}}{2\bar{b}+1} \frac{1}{b_i} \right) / \left(\frac{4(\bar{b}-\bar{\beta})}{(2\bar{b}+1)(2\bar{\beta}+1)} \right) \geq -\frac{1}{2\bar{b}}.$$

But,

$$\begin{aligned} (*) &= \frac{2\bar{\beta}(2\bar{b}+1)b_i - 2\bar{b}(2\bar{\beta}+1)\beta_i}{4b_i\beta_i(\bar{b}-\bar{\beta})} \\ &= \frac{2\bar{\beta}(2\bar{b}+1)(b_i-\beta_i) - 2\beta_i(\bar{b}-\bar{\beta})}{4b_i\beta_i(\bar{b}-\bar{\beta})} \\ &\geq -\frac{1}{2b_i}. \end{aligned}$$

C Proof of lemma 4

Let us introduce a new notation. For all $i \in [\![1; d]\!]$, x and y in $\mathbf{R}^d$, let us denote:

$$[x, y]^{(i)} = (x_1, \ldots, x_{i-1}, 0, y_{i+1}, \ldots, y_d).$$

Let us fix k and l in $\mathcal{Z}_\varepsilon$. We are interrested in the following quatity:

$$b_{k\wedge l}(t) - b_l(t) = \int_{\mathbf{R}^d} K(u) \left(f(t - h^{(k\wedge l)}.u) - f(t - h^{(l)}.u) \right) du.$$

Let us consider the set $J \subset [\![1; d]\!]$ defined by:

$$J = \left\{ i \in [\![1; d]\!]; h_i^{(k)} > h_i^{(l)} \right\}.$$

If $i \in J^c$ then $h_i^{(k\wedge l)} = h_i^{(l)}$. Thus, we denote:

$$\eta_i = \begin{cases} h_i^{(k\wedge l)} = h_i^{(l)} \text{ if } i \in J^c \\ 0 \text{ if } i \in J. \end{cases}$$

Using these notations, we obtain:

$$\begin{aligned} b_{k\wedge l}(t) - b_l(t) &= \int_{\mathbf{R}^d} K(u)\left(f(t - h^{(k\wedge l)}.u) - f(t-\eta.u)\right) du \\ &\quad + \int_{\mathbf{R}^d} K(u)\left(f(t-\eta.u) - f(t - h^{(l)}.u)\right) du. \end{aligned}$$

Thus, it is enough to study quantities of the following form:

$$\int_{\mathbf{R}^d} K(u)\left(f(t-h.u) - f(t-\eta.u)\right) du$$

where $h = h^{(k\wedge l)}$ else $h = h^{(l)}$. In both case we have $h_i = \eta_i$ if $i \in J^c$ and $h_i \le h_i^{(k)}$ if $i \in J$. Here and later we will consider a such bandwidth h.

It is easy to rewrite the following quatity

$$(*) = f(t-h.u) - f(t-\eta.u),$$

using a telescopic sum. We obtain:

$$(*) \;=\; \sum_{i=1}^{d} f_i(-h_i u_i | t - [\eta, h]^{(i)}.u) - f_i(-\eta_i u_i | t - [\eta, h]^{(i)}.u).$$

As $h_i u_i = \eta_i u_i$ if $i \in J^c$, we deduce that indexes belonging to J^c do not contribute to the sum. Finally:

$$(*) = \sum_{i \in J} f_i(-h_i u_i | t - [\eta, h]^{(i)}.u) - f_i(0 | t - [\eta, h]^{(i)}.u).$$

Now, using that f belongs to the anisotropic class $H(\beta, L)$ it is easy to develop the quantity $(*)$ using a Taylor's formula. If we denote $m_i = \lfloor \beta_i \rfloor$, we obtain:

$$\begin{aligned} (*) &= \sum_{i\in J}\sum_{n=1}^{m_i} f_i^{(n)}(0|t-[\eta,h]^{(i)}.u)\frac{(-h_i u_i)^n}{n!} \\ &\quad + \sum_{i\in J}\frac{(-h_i u_i)^{m_i}}{m_i!}\left(f_i^{(m_i)}(\theta_i|t-[\eta,h]^{(i)}.u) - f_i^{(m_i)}(0|t-[\eta,h]^{(i)}.u)\right), \end{aligned}$$

where $|\theta_i| \le h_i |u_i|$.

If we remark that $t - [\eta, h]^{(i)}.u$ does not depend on u_i, using hypothesis $(K4)$ on K and Fubini's theorem, we obtain that, for all $i \in J$ and $n \in [\![1; m_i]\!]$, we have:

$$\int_{\mathbf{R}^d} K(u) f_i^{(n)}(0|t-[\eta,h]^{(i)}.u)\frac{(-h_i u_i)^n}{n!} du = 0.$$

Moreover it is easy to obtain that if $i \in J$, then:

$$\begin{array}{rcl} \left| f_i^{(m_i)}(\theta_i | t - [\eta, h]^{(i)}.u) - f_i^{(m_i)}(0 | t - [\eta, h]^{(i)}.u) \right| & \leq & L|\theta_i|^{\beta_i - m_i} \\ & \leq & L h_i^{\beta_i - m_i} |u_i|^{\beta_i - m_i}. \end{array}$$

Then, we can deduce that:

$$\begin{array}{rcl} \left| \int_{\mathbf{R}^d} K(u) \left(f(t - h.u) - f(t - \eta.u) \right) du \right| & \leq & L \sum_{i \in J} \left(\int_{\mathbf{R}^d} |K(u)| \frac{|u_i|^{\beta_i}}{m_i!} du \right) h_i^{\beta_i} \\ & \leq & L \sum_{i=1}^{d} \lambda_i(\beta) h_i^{\beta_i}. \end{array}$$

Lemma follows.

D Proof of lemma 5

First of all, let us remark that:

$$\forall i, \forall (\beta, L), \quad 1 \leq \frac{h_i(\beta, L, \varepsilon)}{h_i^{(k(\beta, L, \varepsilon))}} \leq 2.$$

Let us calculate:

$$\begin{array}{rcl} B^{\beta, L}(k(\beta, L, \varepsilon)) & = & b^{\beta, L}(h^{(k(\beta, L, \varepsilon))}) \\ & = & L \sum_{i=1}^{d} \lambda_i(\beta) \left(h_i^{(k(\beta, L, \varepsilon))} \right)^{\beta_i} \\ & = & L \sum_{i=1}^{d} \lambda_i(\beta) \left(h_i(\beta, L, \varepsilon) \right)^{\beta_i} \left(\frac{h_i^{(k(\beta, L, \varepsilon))}}{h_i(\beta, L, \varepsilon)} \right)^{\beta_i} \\ & \leq & b^{\beta, L}(h(\beta, L, \varepsilon)). \end{array}$$

On the other hand, it is easy to prove that:

$$b^{\beta, L}(h(\beta, L, \varepsilon)) = \left(\sum_{i=1}^{d} \lambda_i(\beta) \right) s_\varepsilon(h(\beta, L, \varepsilon)).$$

Thus, we obtain:

$$B^{\beta, L}(k(\beta, L, \varepsilon)) \leq (d\lambda^*) s_\varepsilon(h(\beta, L, \varepsilon)).$$

Now, we focus our attention on $s_\varepsilon(h(\beta, L, \varepsilon))$. First, it is easy to prove that:

$$s_\varepsilon(h(\beta, L, \varepsilon)) \leq \frac{\rho_\varepsilon(\beta, L)}{\sqrt{1 + |k(\beta, L, \varepsilon)| \ln 2}} S_\varepsilon(k(\beta, L, \varepsilon)).$$

Next, let us prove that:

$$\frac{\rho_\varepsilon^2(\beta, L)}{1 + |k(\beta, L, \varepsilon)| \ln 2} = \frac{1 + x_\varepsilon(\beta, L)}{1 + |k(\beta, L, \varepsilon)| \ln 2} \leq 2 \vee \frac{\bar{b} + 1}{\bar{b}}.$$

It is known that:

$$1 + |k(\beta, L, \varepsilon)| \ln 2 \geq 1 + x_\varepsilon(\beta, L) - \frac{1}{2\bar{\beta} + 1} \ln(1 + x_\varepsilon(\beta, L)), \tag{13}$$

thus, we obtain that:

$$x_\varepsilon(\beta, L) \leq |k(\beta, L, \varepsilon)| \ln 2 + \frac{1}{2\bar{\beta} + 1} \ln(1 + x_\varepsilon(\beta, L)).$$

This implies in particular that:

$$x_\varepsilon(\beta, L) \leq |k(\beta, L, \varepsilon)| \ln 2 + \sqrt{|k(\beta, L, \varepsilon)| \ln 2}.$$

If $|k(\beta, L, \varepsilon)| \geq 2$ (for example if $\bar{\beta} < \bar{b}/2$), we obtain:

$$x_\varepsilon(\beta, L) \leq 2|k(\beta, L, \varepsilon)| \ln 2,$$

which implies immediatly that:

$$\frac{1 + x_\varepsilon(\beta, L)}{1 + |k(\beta, L, \varepsilon)| \ln 2} \leq 2.$$

On the other hand, when $\bar{\beta} \geq \bar{b}/2$, we obtain from (13) that:

$$\begin{aligned} 1 + |k(\beta, L, \varepsilon)| \ln 2 &\geq 1 + x_\varepsilon(\beta, L) - \frac{1}{\bar{b} + 1} \ln(1 + x_\varepsilon(\beta, L)) \\ &\geq 1 + \frac{\bar{b}}{\bar{b} + 1} x_\varepsilon(\beta, L). \end{aligned}$$

Last inequality implies that:

$$\frac{1 + x_\varepsilon(\beta, L)}{1 + |k(\beta, L, \varepsilon)| \ln 2} \leq \frac{\bar{b} + 1}{\bar{b}}.$$

Lemma follows.

E Proof of lemma 6

First, it is easy to prove that;

$$S_\varepsilon(k(\beta, L, \varepsilon)) \le 2^{d/2} \sqrt{\frac{1 + |k(\beta, L, \varepsilon)| \ln 2}{1 + x_\varepsilon(\beta, L)}} s_\varepsilon(h(\beta, L, \varepsilon)).$$

Next, we know that:

$$1 + |k(\beta, L, \varepsilon)| \ln 2 = 1 + x_\varepsilon(\beta, L) - \frac{1}{2\beta} \ln(1 + x_\varepsilon(\beta, L)) \le 1 + x_\varepsilon(\beta, L).$$

Lemma is proved.